Analysis of the *sltA* (*stzA*) gene and its orthologues in *Aspergillus nidulans* and other filamentous fungi

A thesis submitted in partial fulfilment of the requirements of the University of Wolverhampton for the degree of Doctor of Philosophy

June 2013

Ian James Chilton, BSc (Hons), MSc

This work or any part thereof has not previously been presented in any form to the University or to any other body whether for the purposes of assessment, publication or for any other purpose (unless otherwise indicated). Save for any express acknowledgements, references and/or bibliographies cited in the work, I confirm that the intellectual content of the work is the result of my own efforts and of no other person.

Signature: _____

Date: _____

Abstract

Generation and phenotypic analyses of *stzA* gene deletion strains of *Aspergillus nidulans* implies that *stzA* is allelic to *sltA*, with the encoded transcription factor regulating tolerance to cations, DNA-damaging agents and high arginine concentrations. The similar severe sensitivity of a *sltA1* mutant (GO281) and *stzA* deletion mutants to these stresses indicated that the premature termination codon in *sltA1* represents a total loss-of-function mutation. It was also verified that StzA has no regulatory role in the utilisation of carbon sources. Findings were supported by phenotypic analyses of recombinant progeny resulting from sexual crosses between *sltA1* and *sltA*$^{+}$ strains.

Bioinformatic analysis of genes involved in the osmotic stress response revealed that their promoters were significantly enriched for StzA binding site motifs compared to those of the control group, indicating that StzA may regulate many of these genes that comprise the High Osmolarity Glycerol (HOG) pathway. Although this pathway is activated by fludioxonil, *stzA* deletants and *stzA*$^{+}$ strains showed similar sensitivities to this fungicide. Phenotypic analyses indicate that StzA does not regulate tolerance to sources of oxidative stress, non-ionic osmotic stress or components of the Cell Wall Integrity (CWI) pathway.

A. nidulans StzA appears to have no role in cellulase or xylanase expression as revealed by the results of a dinitrosalicylic acid (DNS) assay. *Trichoderma reesei ace1* deletant and wild-type strains showed similar sensitivities to cations, DNA-damaging agents, arginine, neomycin, acidic and alkaline pH. These results confirm that *A. nidulans* StzA and *T. reesei* Ace1 regulate the distinct phenotypes of abiotic stress tolerance and cellulase and xylanase expression, respectively, despite these two proteins sharing 58% overall amino acid similarity.

All twenty-nine StzA orthologues identified are restricted to filamentous ascomycetes of the Pezizomycotina subphylum and may therefore represent specific and novel antifungal drug targets. The C-termini of StzA proteins are highly variable in both length and sequence, with no apparent conservations in amino acids or predicted secondary structure. This region is considered most likely to influence the divergent functions of StzA proteins. Conservations of individual residues in the N-termini correspond to conserved secondary structure (alpha helices) among StzA proteins, implying shared functions for StzA proteins in this region.

Regulators of two major nitrogen metabolic pathways (CpcA and AreA) may regulate *stzA* expression. Statistically significant putative CpcA binding sites were positionally conserved in 26 out of 29 *stzA* orthologue promoters, indicating an interaction between *stzA* and CpcA, a transcription factor that mediates the cross pathway control of amino acid biosynthesis. *REALALE* sequences, likely to be of retrotransposon origin, containing putative overlapping binding sites for StzA and AreA, were found in eleven *stzA* promoters of the Eurotiomycetes class, indicating an interaction between *stzA* and the global nitrogen metabolite repressor AreA. Intriguingly, *REALALE*-containing promoters identified across the genome of *A. nidulans* were significantly enriched for StzA binding site motifs when compared to a control group of genes. Hence, *REALALE* may have regulatory significance that extends to other *A. nidulans* genes.

Table of Contents

List of Figures

List of Tables

Appendices

Page

Acknowledgements

All work described in this thesis was carried out in the Biosciences Division of the School of Applied Sciences at the University of Wolverhampton.

I would like to express my sincere thanks and appreciation to the members of my supervisory team: Dr. M. P. Whitehead (Director of Studies), Dr. P. Hooley and Dr. D. A. Fincham for their enduring support and guidance. I thank Prof. T. J. Hocking for funding my research.

I am also extremely grateful to those who have provided me with technical support: Andy Brook, Ann Dawson, Balbir Bains, Clare Murcott, Nick Skidmore and Colleen Delaney. Additionally, I thank Dr. Malcolm Inman for his excellent photographic work and Dr. Alison McCrea for her valuable advice with statistical analysis.

I would like to thank Michael L. Nielsen and Uffe H. Mortensen (Technical University of Denmark) for supplying the plasmid pDEL1 and for their useful advice on gene deletion. Thanks to Nino Aro and Merja Pentillä (VTT Biotechnology, Finland) for supplying the *Trichoderma reesei ace1* deletion and wild-type strains. Thanks also to Dr. Nicholas Helps and his team at the Dundee Sequencing Centre for their DNA sequencing work.

I wish the researchers with whom I have shared the labs the very best in their future endeavours.

Finally, I especially thank my mother and father for their unwavering love and support throughout my studies.

Abbreviations

aa	Amino acid
ADP	Adenosine diphosphate
AGT	O^6-alkylguanine DNA alkyltransferase
AIDS	Acquired immune deficiency syndrome
AspGD	*Aspergillus* Genomes Database
ATM	Mutated in ataxia telangiectasia
ATP	Adenosine triphosphate
ATR	ATM-Rad3-related
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BSA	Bovine serum albumin
CADRE	Central *Aspergillus* Data REpository
CCR	Carbon catabolite repression
CDD	Conserved Domain Database
cDNA	Complementary DNA
ChIP-chip	Chromatin immunoprecipitation and 'chip' (microarray) analysis
ChIP-seq	Chromatin immunoprecipitation and sequencing
CWI	Cell wall integrity
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease
DNS	Dinitrosalicylic acid
dNTP	Dinucleotide triphosphate
DSBs	Double-strand breaks
DWA	Distilled water agar
EB	Extraction buffer
EDTA	Ethylene diamine tetra-acetic acid
EMS	Ethyl methanesulfonate
EST	Expressed sequence tag
***et al*.**	*et alia*
FFGED	Filamentous Fungal Gene Expression Database
FGSC	Fungal Genetics Stock Centre
FOA	Fluoro-orotic acid
g	Gram
g	Gravitational force
GFD	Glycerol-3-phosphate dehydrogenase
GFP	Green fluorescent protein
GLD	Glycerol dehydrogenase
GPP	Glycerol-3-phosphate phosphatase
GTP	Guanosine triphosphate
HIV	Human immunodeficiency virus
HK	Histidine kinase
HOG	High osmolarity glycerol
HPt	Histidine-containing phosphotransfer intermediate
HR	Homologous recombination
kb	Kilobase
λ	Lambda
L	Litre
LB	Luria Bertini broth
LBA	Luria Bertini agar
LTR	Long terminal repeat
M	Molar
MAPK	Mitogen-activated protein kinase
ME	Malt extract
MEA	Malt extract agar
mg	Milligram
min	Minute
MIPS	Munich Information Center for Protein Sequences

ml	Millilitre
mM	Millimolar
MM	Minimal media
MMS	Methyl methanesulfonate
MNNG	*N*-Methyl-*N*'-Nitro-*N*-Nitrosoguanidine
MOPS	3-(N-morpholino)-propanesulfonic acid
MPT	Methylphosphotriester
mRNA	Messenger RNA
n	Nucleotide
N	Total sample size
NCBI	National Center for Biotechnology Information
NER	Nucleotide excision repair
ng	Nanogram
NHEJ	Non-homologous end-joining
NLS	Nuclear localisation sequence/signal
NMD	Nonsense mediated decay
NMR	Nitrogen metabolite repression
4NQO	4-nitroquinoline oxide
nt	Nucleotide
ºC	Degree Celsius
OD	Optical density
ORF	Open reading frame
P	Associated probability
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PEPCK	Phosphoenolpyruvate carboxykinase
PFK	Phosphofructokinase
PIKK	Phosphatidylinositol-3 kinase-related kinase
PRR	Post-replication repair
psi	Pounds per square inch
PTC	Premature termination codon
qPCR	Quantitative polymerase chain reaction
RACE	Rapid amplification of cDNA ends
REALALE	*REAL*-associated little element
RNA	Ribonucleic acid
RNAse	Ribonuclease
rpm	Revolutions per minute
RR	Response regulator
s	Second
SDS	Sodium dodecyl sulphate
SDW	Sterile distilled water
SE	Standard error
Taq	*Thermus aquaticus*
TBE	Tris borate EDTA
TCA	Tricarboxylic acid
TE	Tris EDTA
TF	Transcription factor
TLC	Thin layer chromotography
T_m	Melting temperature
Tris	Tris (hydroxymethyl) aminomethane
U	Units
µg	Microgram
µl	Microlitre
UV	Ultraviolet
V	Volt
χ^2	Chi-square

Amino acid abbreviations

Amino acid	IUPAC code	
	3-letter	1-letter
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic Acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V
Unspecified	Xaa	X

Nucleic acid abbreviations

Nucleic acid	IUPAC code
Adenine	A
Cytosine	C
Guanine	G
Thymine (or Uracil)	T (or U)
A or G	R
C or T	Y
G or C	S
A or T	W
G or T	K
A or C	M
C or G or T	B
A or G or T	D
A or C or T	H
A or C or G	V
Any base	N

Chapter 1

Introduction

1.1 The genus *Aspergillus*

Approximately 70,000 fungal species have been described and the estimated total is 1.5 million (Lutzoni *et al*., 2004). The ascomycete genus *Aspergillus* includes over 185 species found worldwide with a diverse range of lifestyles in soil, decaying vegetation, household dust, building materials, food and water (Warris and Verweij, 2005). A large number of *Aspergillus* species are of biomedical and industrial significance, but many species within this genus are also extremely harmful as well as beneficial to humans. For example, this genus includes *Aspergillus nidulans* (a model organism for genetics and cell biology), *Aspergillus fumigatus* (a life-threatening pathogen responsible for aspergillosis in humans and other animals), *Aspergillus flavus* (a serious pathogen of both plants and animals) and *Aspergillus oryzae* (a source of many enzymes used in the food industry; reviewed in Machida and Gomi, 2010).

Genome projects for several *Aspergillus* species are now complete and reveal considerable variation among its members (Wortman *et al*., 2009). The profoundly differing lifestyles exhibited by each of this growing set of *Aspergillu*s species for which genome sequences are available, coupled with the varying degrees of evolutionary affinity shared by their genomes, make *Aspergillus* a model clade to address fundamental questions in functional and comparative genomics. Comparisons of pairs of these genomes may pinpoint genes responsible for adaptation to their environment and ultimately account for their divergence into separate species (Galagan *et al*., 2005; Rokas *et al*., 2007; Wortman *et al*., 2009). Recombination hotspots, often close to the telomeres, are associated with rearrangements and inversions that are the likely origins for speciation events (Galagan *et al*., 2005).

A comparative analysis of the genomes of *A. nidulans* with *A. fumigatus* and *A. oryzae* revealed that these organisms have diverged significantly, despite the fact these organisms are members of the same genus and are sufficiently related that orthologues can be identified for the majority of genes. For example, the number of predicted protein coding genes in the *A. nidulans* (9,541) and *A. fumigatus* (9,926) genomes are similar, whereas *A. oryzae* is predicted to have over 4,000 additional genes (14,063). Furthermore, proteome comparisons revealed an average amino acid identity of less than 70% between each species pair, suggesting that they are as evolutionary distant from each other as humans are from fish (Galagan *et al*., 2005).

1.2 The model organism *Aspergillus nidulans*

The filamentous ascomycete *Aspergillus nidulans* has been used as a model eukaryotic organism for more than sixty years. This organism has proved an attractive model system for studying the effects of changes in environmental conditions on fungal growth due to its rapid life cycle, ease of manipulation and amenability to physiological, genetic and molecular techniques (reviewed in Archer and Dyer, 2004). Furthermore, *A. nidulans* has a range of well-characterised single-gene mutants whose behaviour can be compared directly with wild-type strains to enable the roles of particular genes to be determined (Chaveroche *et al*., 2000). Mutant *Aspergillus* strains, isolated over many years, are available at the Fungal Genetics Stock Centre (FGSC; www.fgsc.net). The completed genome sequence of *A. nidulans* will contribute greatly to our understanding of the physiology of this organism at the molecular level (Galagan *et al.*, 2005). The annotated genomes of *A. nidulans* and other *Aspergilli* are available at the Central *Aspergillus* Data REpository database (C*A*DRE; http://www.cadre-genomes.org.uk/index.html; Mabey *et al*., 2004) and the *Aspergillus* Genome Database (AspGD; http://www.aspgd.org/; Arnaud *et al*., 2010). These resources are intended to provide well-annotated reference genomes, with detailed gene and protein information, for the *Aspergillus* clade of organisms. Enhanced understanding of the physiology of *A. nidulans* will be aided by Polymerase Chain Reaction (PCR)-based gene targeting techniques developed for *A. nidulans* by Nielsen *et al*. (2006), Nayak *et al*. (2006) and Szewczyk *et al*. (2006) and subsequently applied to other filamentous fungal species (Krappmann, 2007; Larrondo *et al*., 2009).

1.3 Environmental stress

Stress has been defined as any factor that reduces the growth rate below the optimum for an organism (Mager and De Kruijff, 1995). A fungus growing in a saprophytic environment will encounter a wide range of stresses, including osmotic stress, DNA-damaging agents, nutrient limitation, oxidative stress, secretion stress and adverse temperature and pH (Edwards *et al*., 1998). Hence, it is not surprising that fungi growing in highly challenging environments have developed integrated control systems for withstanding these stresses. It is clearly of considerable importance to understand the mechanisms enabling adaptation to these stresses, as fungi are some of the most serious pathogens of plants, causing serious food spoilage, and several species cause life-threatening infections of animals (Archer and Dyer, 2004).

1.3.1 The general stress response in yeasts

A general stress response can be defined as a stress response in which the same set of genes is induced in response to a variety of different stresses (Enjalbert *et al*., 2003). This is different from specific stress responses where specific genes are induced by specific stresses (e.g. induction of heat shock proteins in response to heat shock or induction of compatible solutes in response to salt stress). The general stress response allows cells that have been challenged with a mild stress to acquire tolerance to a stronger stress that may be different from the first stress encountered (Lewis *et al*., 1995).

The general stress response in *S. cerevisae* involves the transcription factors (TFs) Msn2p and Msn4p, which bind to the stress responsive element (STRE) sequence 5'-CCCCT in the promoters of target genes (Martínez-Pastor *et al*., 1996). This response occurs to diverse types of stress, including heat shock, oxidative and reductive stress, osmotic shock, nutrient starvation, DNA damage and extreme pH (Gasch *et al*., 2000). Characteristics of the general stress response include accumulation of glycogen and trehalose, which function as carbohydrate reserves. Groups of genes induced by the general stress response in *S. cerevisae* include those involved in carbohydrate metabolism, defence against reactive oxygen species, protein metabolism, intracellular signalling and DNA-damage defence and repair (Gasch, 2007).

Other yeasts have been analysed for the presence of a general stress response similar to that of *S. cerevisae.* Such a stress response has been identified in *Schizosaccharomyces pombe* (Chen *et al*., 2003). It is, however, apparent that regulation of the general stress response has diverged significantly between *S. cerevisiae* and *S. pombe*, as these species use different TFs to co-ordinate the response (Sty1-Atf1 in *S. pombe* and Msn2/4p and Rpd3p in *S. cerevisae*). Although *S. cerevisae* and *S. pombe* exhibit strong general stress responses, there does not appear to be a general stress response in *Candida albicans* (Enjalbert *et al*., 2003). In *C. albicans*, there was no evidence for induction of a common set of genes in response to temperature, osmotic or oxidative stresses. Furthermore, neither glycogen nor trehalose, known to be hallmarks of the general stress response, was induced in response to mild stress, unlike the situation in *S. cerevisiae*. This indicates that the general stress response was lost in *C. albicans* after it split from the *S. cerevisiae* lineage. It is possible, however, that *C. albicans* mounts a general stress-like response in limited conditions or in a specific morphological state (Gasch, 2007).

1.3.2 The HOG pathway in *Saccharomyces cerevisiae*

S. *cerevisiae* reponds to environmental changes by activating mitogen-activated protein kinase (MAPK) signalling pathways. During this process, changes are detected and signals are transduced into appropriate cellular responses to enable cells to adapt to the prevailing environmental conditions. In *S. cerevisiae*, the four MAPKs: Hog1, Fus3, Kss1 and Slt2 are activated in response to osmolarity, mating pheromones, starvation and cell wall damage, respectively (reviewed in Gustin *et al*., 1998; Qi and Elion, 2005).

The high osmolarity glycerol (HOG) MAPK pathway is essential for yeast survival in a high osmolarity environment. This pathway in *S. cerevisiae* has become a model for the study of MAPK signalling systems (Figure 1.1; reviewed in Hohmann, 2009). The key component of the HOG pathway is Hog1 MAPK, which is activated during hyperosmotic stress by phosphorylation of conserved threonine and tyrosine residues, and occurs by two independent upstream mechanisms that converge on Pbs2 MAPKK. Activated Hog1 enters the nucleus and activates the transcription of genes involved in the osmotic stress response.

Two transmembrane osmosensors, Sln1 and Sho1, can independently activate Hog1 *via* the Sln1 and Sho1 branches of the HOG pathway (Hohmann, 2009). Under hyperosmotic conditions, Sln1 responds to changes in cellular turgor pressure by altering its hexokinase (HK) domain to an inactive conformation, resulting in cessation of phosphate transfer *via* the Sln1-Ypd1-Ssk1 phospho-relay reaction. Dephosphorylated Ssk1 binds to the regulatory domain of the Ssk2 and Ssk22 MAPKKKs, causing them to autophosphorylate and activate themselves. Ssk2 and Ssk22 phosphorylate and activate Pbs2, which subsequently activates Hog1 by phosphorylating two of its residues (Thr174 and Tyr176).

The Sho1 branch of the HOG pathway is regulated by Msb2 and Hkr1, two mucin-like transmembrane sensors (Tatebayashi *et al*., 2007). Pbs2 serves as MAPKK and scaffold for the Sho1 branch, but many of the molecular details of the activation mechanism remain unknown (Chen and Thorner, 2007). It is thought that Pbs2 carries Ste11 MAPKKK to the vicinity of Ste20 and Cla4 kinases, which are associated with the Cdc42 G-protein at the plasma membrane. Ste11 is activated by phosphorylation by Ste20 and/or Cla4, and Ste11 subsequently phosphorylates Pbs2, which in turn phosphorylates and activates Hog1 (Hohmann, 2009). Various protein phosphatases control the phosphorylation state of MAPK Hog1 and include the phospho-tyrosine phosphatases Ptp2 and Ptp3 and the phospho-threonine phosphatases Ptc1, Ptc2 and

Ptc3. Phosphorylation of Hog1 is accompanied by import of Hog1 into the nucleus, where it activates Hot1 and other HOG-mediated TFs. Hog1 also has some target genes in the cytosol that remain to be identified (Saito and Tatebayashi, 2004).

It is not clear why the Hog1 pathway has two separate branches, as either branch alone can activate Hog1 in response to hyperosmotic stress. Under moderate osmotic shock, both the Sho1 and Sln1 branches are necessary for the adaptation process, but under severe osmotic shock the Sln1 branch has a more dominant role in mediating the osmotic stress response and remains fully activated, even in the absence of the Sho1 branch (O'Rourke and Herskowitz, 2004; Parmar *et al*., 2009). The greater importance of the Sho1 branch may explain its greater evolutionary conservation when compared with the Sho1 branch in yeasts and other fungi (Parmar *et al*., 2009). The fact that the Sho1 branch is not connected to the MAPK cascade of Hog1 in a number of fungi, including *A. nidulans*, would appear to further demonstrate the redundancy of this pathway in the osmotic stress response (Furukawa *et al*., 2005). The Sho1 branch may, however, co-ordinate signals between Hog1 and other MAPK pathways. An overlap in communication between signal transduction pathways is evident, as the Sho1 branch of the HOG pathway shares protein kinases with the Fus3 and Kss1 pathways, and all the MAPK pathways share protein phosphatases (Hohmann, 2009). Further elucidation of the interactions within the yeast MAPK network is necessary (Chen and Thorner, 2007).

The HOG pathway performs the important role of controlling glycerol accumulation. Glycerol is important as a compatible solute for osmoregulation and for redox-balancing. Hog1 controls expression of the genes encoding glycerol-3-phosphate dehydrogenase (Gpd1) and glycerol-3-phosphatase (Gpp1 and Gpp2; Rep *et al*., 2000). Hog1 is also involved in the expression of the Stl1 active glycerol uptake system (Rep *et al*., 2000) and controls the activity of the glycerol export channel Fps1 to prevent glycerol leakage (Tamás *et al*., 2003). Additionally, Hog1 appears to control the activity of the enzyme phosphofructokinase (PFK), which yields the glycolytic activator fructose 2,6-bisphosphate to enhance glycerol production (Hohmann, 2009).

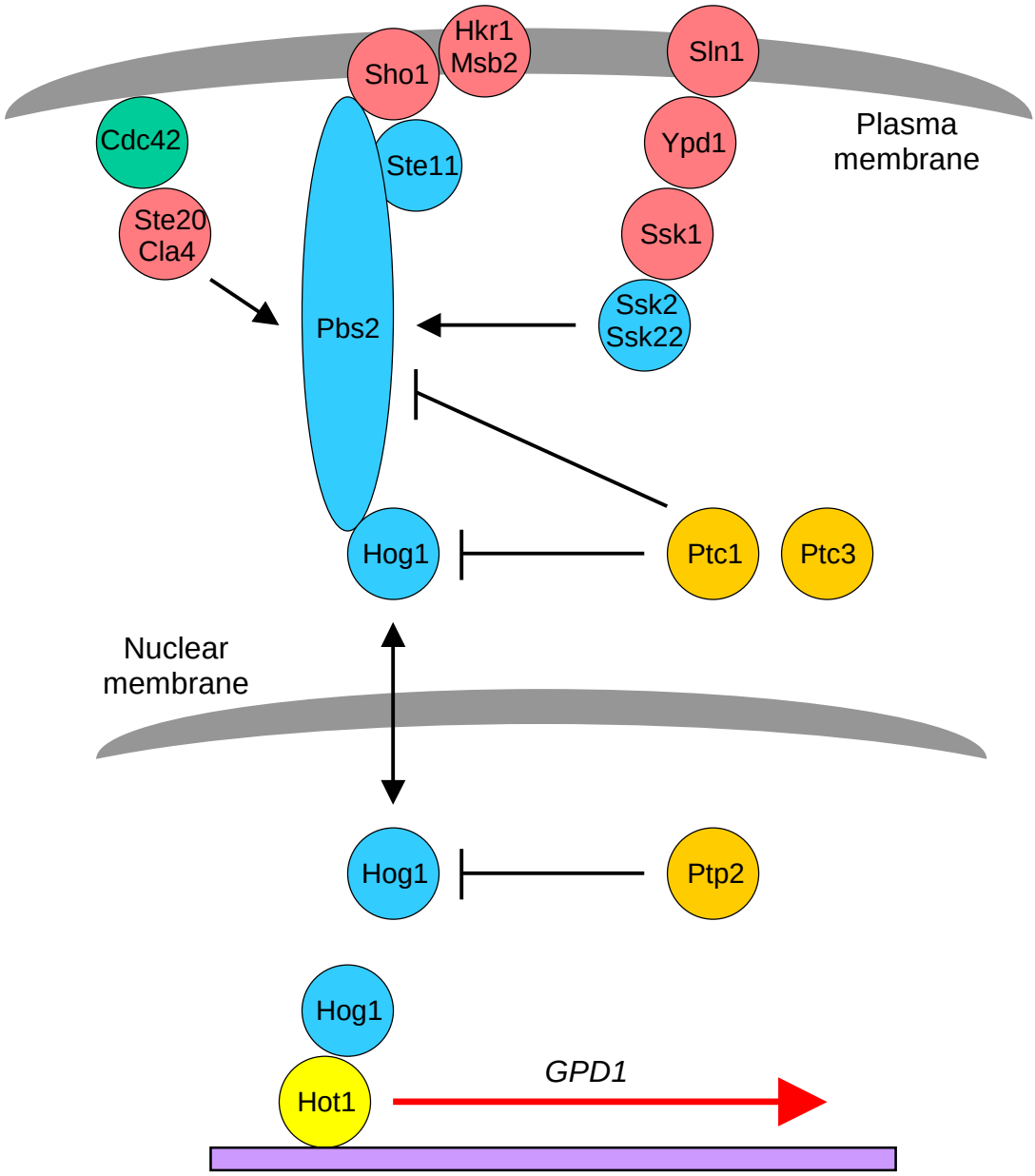


Figure 1.1. The two branches (Sln1 and Sho1) of the HOG pathway. Adapted from Hohmann (2009). The Sln1 and Sho1 branches of the HOG pathway converge on Pbs2 to activate Hog1, which accumulates in the nucleus under conditions of osmotic stress. Within this system, Ste11, Ssk2 and Ssk22 are MAPKKKs, Pbs2 is a MAPKK and Hog1 is the MAPK. Membrane-localised sensors are shown in red, protein kinases in blue, protein phosphatases in orange, the Cdc42 GTPase in green and the Hot1 TF in yellow.

1.3.3 Osmotolerance mechanisms in filamentous fungi

Among eukaryotes, it is within fungi that responses to osmotic stress are best understood (reviewed in Hagiwara *et al*., 2009; Miskei *et al*., 2009). The principal soil osmoticant is NaCl. High extracellular NaCl causes efflux of water from the fungal cell by osmosis, leading to dehydration. However, the direct uptake of NaCl to maintain cellular water content is an unfavourable strategy as high intracellular Na^+ has a detrimental effect on cellular metabolic processes, due to its toxic effect on cytosolic enzymes (Blomberg and Adler, 1992). Enzymes of both terrestrial and marine species of fungi and plants are generally salt-sensitive, with enzyme inhibition occurring above 100–200 mM NaCl, which is approximately 20–40% of the concentration of sea water (Blomberg and Adler, 1992). This is in contrast to halophilic archea, which not only tolerate salt but prefer high salt concentrations (2–6 M KCl). Instead of excluding salt, these organisms have the same ionic strength inside their cells as outside, so possess proteins that have adapted to remain folded and functional in a hypersaline environment. These proteins have an abundance of negatively charged residues combined with a low frequency of lysines (Tadeo *et al*., 2009). Fungi have adapted to survive in saline environments by excluding toxic concentrations of ions from the cytosol and by synthesising non-toxic osmotically active compounds in order to retain water. The ability to withstand osmotic stress is particularly important in marine fungi, including the yeast *Debaryomyces hansenii* and the filamentous fungi *Dendryphiella salina* and *Dendryphiella arenaria*, to prevent continual water loss to the surrounding saline environment (~0.5 M NaCl; Clipson and Jennings, 1992; Edwards *et al*., 1998).

In environments with high Na^+ or K^+, there is almost inevitable influx of these cations into fungal cells due to their highly negative membrane potential (Rodriguez-Navarro *et al*., 1986). Hence, this cation influx needs to be balanced with mechanisms that achieve Na^+ or K^+ efflux in order to protect enzymes. Na^+/H^+ antiporters achieve the exclusion of Na^+ from the cytosol in both marine and terrestrial fungi (Clement *et al*., 1999). These ion exchangers are present in the fungal cell membrane and the membranes of cellular compartments, to exclude Na^+ from the cytosol by pumping Na^+ out of the cell or by sequestering Na^+ into vacuoles. This process uses the energy created by the transport of H^+ down their concentration gradient to transport Na^+ up their concentration gradient. The vacuole plays an important role in salt tolerance by sequestering Na^+ to compartmentalise this toxic ion from the salt-sensitive metabolically active cytoplasm. In yeast, the Na^+/H^+ exchanger, Nhx1, contributes to osmotolerance by sequestering and compartmentalising Na^+ in the vacuole (Nass and Rao, 1999) and a similar mechanism operates in other fungi, including *A.*

nidulans (Redkar *et al*., 1995). In *A. nidulans*, selective uptake of K^+ over Na^+ has been observed in conditions of osmotic stress when both these ions are abundant in the extracellular environment (Beever and Laracy, 1986). Under these circumstances, K^+ is used for osmotic adjustment, as high intracellular concentrations of K^+ can be tolerated by the cell, while the accumulation of cytotoxic levels of Na^+ is prevented.

Na^+/H^+ and K^+/H^+ antiporters cannot be relied upon in environments with high pH because in these conditions the membrane potential is too low to drive their function. To meet the challenge of surviving salt stress in alkaline environments, fungi are furnished with Na^+- and K^+-efflux ATPases in their cell membranes to actively pump Na^+ and K^+ out of the cell without involvement of protons (Benito *et al*., 2002). Such ATPases would seem to be especially relevant to osmoadaptation in fungi that complete their life cycles in the alkaline (~pH 8.5), salty (~0.5 M) environment of seawater. However, it is likely that plasma membrane cation-efflux ATPases exist in all fungi, as genes encoding homologues of Na^+ and K^+ ATPases have been cloned and characterised from diverse terrestrial yeasts and filamentous fungi, including *A. nidulans* and *Neurospora crassa* (Benito *et al*., 2002), in addition to the marine hyphomycete *D*. *salina* (Davies *et al*., 1990). The cell wall may also contribute greatly to osmotolerance in (marine) filamentous fungi as this structure was found to contain higher toxic concentrations of Na^+ and K^+ compared to the rest of the cell in *D*. *salina* (Clipson *et al*., 1989). Considering that the cell wall contributes between 36% and 49% of the total cell volume, this structure has the potential to be efficiently utilised in the prevention of K^+ and Na^+ ions from passing through the plasma membrane into the cytoplasm.

The exclusion of toxic concentrations of ions from the cytosol protects salt-sensitive proteins, but this strategy does not favour water retention in conditions of osmotic stress. Fungi must control their cytoplasmic osmotic pressure by synthesising and accumulating metabolically inactive osmolytes, known as compatible solutes, which include polyhydric alcohols (polyols), proline and trehalose (Beever and Laracy, 1986). The function of these compounds is to cause a net flux of water into the cell to favour the retention of cellular water. Water influx is necessary for hyphal growth and hyphal turgor, allowing the cell to grow normally. In addition, hyphal turgor allows growth of the fungus through soils and the solid tissues of plants and animals (Davis *et al*., 2000a). The polyols glycerol, erythritol, arabinitol and mannitol are the compatible solutes that have been shown to be present in the highest concentrations in fungi, including *S. cerevisiae*, *A. nidulans*, *D*. *hansenii* and *D*. *salina* (Beever and Laracy, 1986; Redkar *et al*., 1995). In many fungi, glycerol plays a major

role in regulating the osmotic potential of the cell. It has been shown that glycerol is the major polyol synthesised in *S. cerevisiae* (Blomberg and Adler, 1992) and *A. nidulans* (Redkar *et al*., 1995) in response to increasing salt concentrations. Conversely, lack of glycerol accumulation in these species resulted in osmosensitivity. Glycerol is understood to be the main compatible solute produced by fungi for two main reasons. Firstly, glycerol is the smallest polyol and can therefore be quickly synthesised, making it more efficient for coping with osmotic stress. Secondly, glycerol has been shown to have a stabilising effect on enzymes (Witteveen and Visser, 1995). Additionally, it has been suggested that glycerol is important for redox balancing; i.e. the removal of excess NADH during growth at reduced water activity (Ruijter *et al*., 2003). The synthesis of glycerol under osmotic stress conditions in *A. nidulans* is mainly dependent on NADP-dependent glycerol dehydrogenase (GLD) activity rather than glycerol-3-phosphate dehydrogenase (GFD) activity, in contrast to the situation in *S. cerevisiae* where the reverse was observed. Thus, in *A. nidulans*, dihydroxyacetone phosphate is converted firstly to dihydroxyacetone and then to glycerol by glycerol dehydrogenase (Figure 1.2; De Vries *et al*., 2003). In *Aspergillus niger*, mannitol is essential for the protection of spores against cell damage caused by salt stress, and trehalose is necessary for long-term spore survival and spore germination in response to salt stress (Ruijter *et al*., 2003).

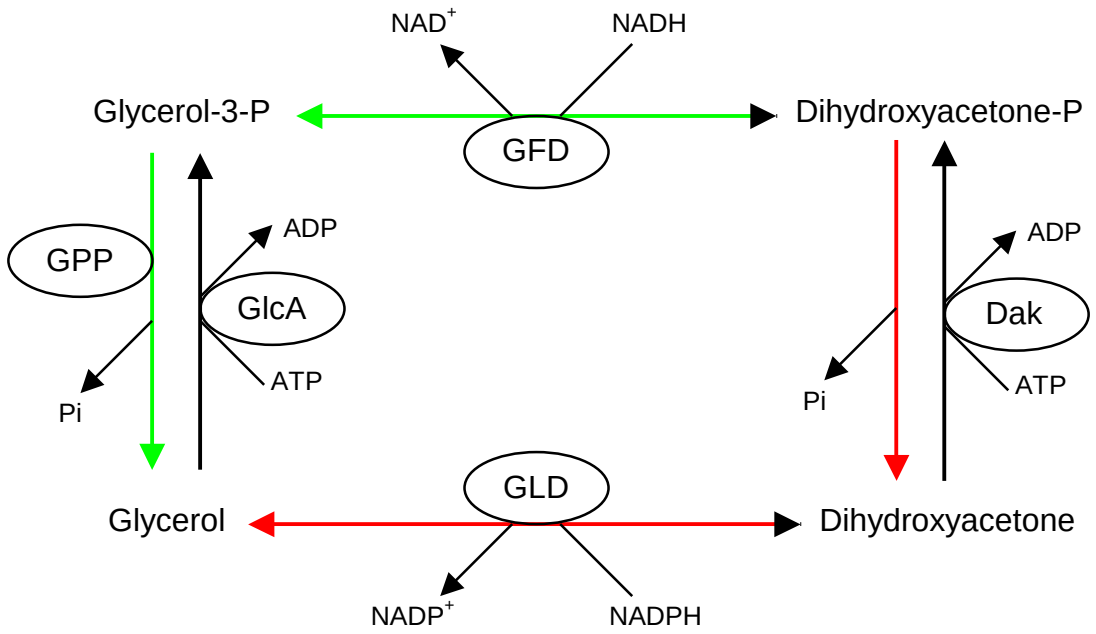


Figure 1.2. Biosynthetic pathways for the conversion of dihydroxyacetone to glycerol in fungi. Adapted from De Vries *et al*. (2003). Glycerol is the major polyol synthesised in response to osmotic stress in both *S. cerevisiae* and *A. nidulans*, but is synthesised *via* two different pathways from dihydroxyacetone phosphate. In *S. cerevisiae*, glycerol is synthesised from *via* glycerol-3-phosphate, whereas in *A. nidulans* it is synthesised predominantly *via* dihydroxyacetone. These pathways are indicated by green and red arrows, respectively. GFD, glycerol-3-phosphate dehydrogenase; GPP, glycerol-3-phosphate phosphatase; GLD, glycerol dehydrogenase; GlcA, glycerol kinase; Dak, dihydroxyacetone kinase.

1.3.4 DNA repair mechanisms in *A. nidulans*

The DNA damage response in *A. nidulans* ensures maintenance of genome integrity during cellular reproduction (Goldman and Kafer, 2004). Types of DNA damage include single- and double-strand breaks, base damage and DNA-protein crosslinks, which may have various causes that include ultraviolet (UV) irradiation, X-rays and alkylating agents (Goldman *et al*., 2002). Such damage can interfere with the function of replication forks and lead to the accumulation of secondary DNA mutations that result in genomic instability. The pathways involved in the response to DNA damage either repair the damage or allow it to be tolerated. Goldman and Kafer (2004) have reviewed the DNA repair genes of *A. nidulans* and matched them to their human orthologues. In contrast to humans and other eukaryotic model organisms, the DNA damage response in filamentous fungi is less well understood (Goldman *et al*., 2002; Semighini *et al*., 2005). It is, however, known that the same DNA repair pathways operate in *A. nidulans*, which include photoreactivation, nucleotide excision repair, postreplication repair, recombination repair, non-homologous end-joining, cell cycle arrest and apoptosis (Goldman and Kafer, 2004).

Photoreactivation is a mechanism that repairs DNA damage resulting from UV irradiation, the major source of genotypic damage in *A. nidulans* and other terrestrial organisms. This mechanism uses light-dependent DNA photolyase to break the UV bonds that result from the two types of genotoxic lesions produced by UV irradiation: cyclobutane pyrimidine dimers and 6-4 pyrimidine–pyrrolidine photoproducts (reviewed in Brettel and Byrdin, 2010).

Nucleotide excision repair (NER) is used to repair DNA damage that escapes the mechanism of photoreactivation, and additionally removes bulky base additions that distort the DNA helix (reviewed in Hatakeyama *et al*., 1998). Damaged DNA strands are cut at the 5’ end of the lesion by Rad10 endonuclease and at the 3’ end by Rad2 endonuclease. The strands are unwound by a DNA helicase before the gaps are filled in by a DNA polymerase that uses the undamaged strand as a template. This newly synthesised DNA is ligated to the sites of incision.

The postreplication repair (PRR) system is a network of pathways that provides cells with mechanisms to tolerate DNA damage rather than to repair it (Goldman and Kafer, 2004). PRR is used to overcome stalled replication forks caused by UV photoproducts that escape the repair mechanisms of photoreactivation and

NER. Specialised DNA polymerases can insert bases opposite a lesion in the replication fork to allow replication to proceed beyond the block. However, this occurs at the expense of a mutation since the polymerases do not incorporate bases according to Watson-Crick base pairing rules. A central role in PRR is played by the conserved Rad6/Rad18 complex, which possesses both ATPase and ubiquitin-conjugating activity, and is thought to bind to single-stranded DNA at blocked replication forks and promotes DNA damage tolerance in downstream regions (Cho *et al*., 2003).

Recombination repair is a DNA repair process that occurs when double-strand breaks (DSBs) occur in a chromosome (Chu, 1997). This process uses gene conversion, whereby sequences from a sister chromatid or a homologous chromosome are used as a template for DNA repair. In *A. nidulans*, UvsC is likely to be required for strand exchange between homologous sequences (van Heemst *et al*., 1997). This form of DSB repair is rarely accompanied by crossing over, unlike meiotic recombination, making this form of repair largely error-free. However, when homologous sequences are not available as a template (during the G1 phase of a haploid cell cycle), non-homologous end-joining (NHEJ) is used to ligate broken strands: a process that is frequently error-prone (Chu, 1997). The Ku70/Ku80-protein complex and the XRCC4-DNA ligase IV complex are essential components of the NHEJ pathway (Nayak *et al*., 2006; Shrivastav *et al*., 2008).

The arrest of the cell cycle is an important aspect of the DNA damage response because it allows time for repair of the genotoxic lesions, reducing the probability that damaged chromosomes will be replicated or segregated. The DNA damage checkpoint blocks entry into mitosis. This process in *A. nidulans*, like other eukaryotes, involves maintaining phosphorylation of the Cdc2 protein kinase orthologue (NimX) at the key Y15 tyrosine residue, which is controlled by the Wee1 and Cdc25 orthologues, AnkA and NimT, respectively (Kraus and Harris, 2001). The *scaA* gene of *A. nidulans* is a novel gene that is likely to regulate recombination and DNA checkpoint functions (Fagundes *et al*., 2003). How these mechanisms are influenced by DNA damage remains to be elucidated.

Apoptosis may have a significant role to play in dealing with irreparable DNA damage in *A. nidulans* and other filamentous fungi (Goldman *et al*., 2002). This is feasible since multicellular organisms, unlike yeasts, could use programmed cell death to ensure that severely damaged genomic DNA does not contribute to future generations. However, the molecular mechanisms underlying hyphal cell death have yet to be

elucidated, but an apoptosis-inducing factor (AifA) has been identified in *A. nidulans* (Dinamarco *et al*., 2010).

Little is known about the global regulation of the filamentous fungal DNA damage response, but it is most likely regulated in a similar manner to that of yeast and human cells (Goldman *et al*., 2002; Semighini *et al*., 2005). Such regulation involves a member of the ATM/ATR (ataxia-telangiectasia mutated/ATM and Rad3-related) subfamily of PIKKs (phosphatidylinositol-3 kinase-related kinases), which is activated by sensors that detect the presence of DNA damage and maintains genome integrity by phosphorylating multiple target proteins. The orthologue of this key member of the ATM/ATR subfamily of PIKKs has been identified in *A. nidulans* (UvsB) as well as the PIKK regulatory subunit Rad26 (UvsD; De Souza *et al*., 1999). Additionally, a RecO helicase, which allows recovery from the DNA damage response appears to be essential in *A. nidulans*, yeast and mammals (Hoffmann and Harris, 2001). Analysis of an *A. nidulans* strain exposed to DNA-damaging agents has indicated that the global regulation of the DNA damage response in filamentous fungi is likely to involve activation of multiple pathways acting in parallel to mediate DNA repair (Malavazi *et al*., 2006). The availability of the complete *A. nidulans* genome sequence is likely to facilitate DNA repair studies that will allow the complexity of this regulation to be unravelled by transcriptional profiling methods such as microarrays and proteomics (Goldman and Kafer, 2004).

1.3.5 Transcriptional regulation of stress responses in *A. nidulans*

Transcriptional regulation of stress responses in fungi contributes to the homeostasis necessary for their survival. In yeasts and filamentous fungi, like other eukaryotes, the most common signal transduction mechanisms involved in cellular responses to environmental signals involve histidine-to-aspartate (His–Asp) phosphorelay systems that activate mitogen-activated protein kinase (MAPK) pathways (reviewed in Qi and Elion, 2005). His–Asp pathways consist of a histidine kinase (HK), a response regulator (RR) and a histidine-containing phosphotransfer intermediate (HPt). MAPK cascades then transduce signals to the nucleus to effect transcriptional reponses, allowing adaptation to the environmental changes encountered. In *S. cerevisiae*, the best documented stress-activated MAPK cascade is the HOG (high-osmolarity glycerol) pathway (Section 1.3.2), which is mainly involved in responses to osmotic stress, but also to oxidative stress and cell cycle regulation (reviewed in Hohmann, 2009).

The molecular mechanisms of stress signalling in filamentous fungi are less well understood. In *A. nidulans*, a model for His–Asp phosphorelay signalling and HogA MAPK cascade activation has been proposed (Figure 1.3; Hagiwara *et al*., 2009). This pathway, involving SskA, HogA (SakA) and AtfA, regulates transcriptional responses to osmotic stress, oxidative stress and the fungicide fludioxonil. In the phosphorelay system of the proposed model, NikA (HK) serves as a sensor and transmits the signal to SskA and SrrA (RRs) *via* YpdA (HPt). It has been shown that AtfA is permanently localised in the nucleus, but HogA only accumulates in the nucleus in response to osmotic or oxidative stress signals, where it interacts with AtfA. This interaction leads to transcriptional up-regulation or down-regulation of genes involved in responses to osmotic stress, fludioxonil (Hagiwara *et al*., 2009) and oxidative stress (Lara-Rojas *et al*., 2011). Many genes up-regulated in response to osmotic stress and fludioxonil include those with putative functions in sugar or polyol transport and metabolism, particularly trehalose. Trehalose biosynthesis is important for the acquisition of stress tolerance in both *A. nidulans (*Fillenger *et al*., 2001) and *A. fumigatus* (Al-Bader *et al*., 2010), including osmotic stress, oxidative stress, heat stress and starvation. The precise role of the stress response regulator SrrA remains unclear. Furthermore, *A. nidulans* has fourteen HKs additional to NikA that have only partially been characterised, so it is quite possible that these may also serve as stress sensors (Suzuki *et al*., 2008; Hagiwara *et al*., 2009). Although AtfA regulates different types of stress responses in *A. nidulans* similarly to Atf1 in *S. pombe*, which has a general stress response, it is unclear whether HogA–AtfA constitutes a general stress response mechanism in *A. nidulans* (Balázs *et al*., 2010; Lara-Rojas *et al*., 2011). Intriguingly, the model for osmotic stress signal transduction in *A. nidulans* may also be applicable to other filamentous fungi, particularly *Aspergillus* species, which contain orthologues of stress signalling components that are widely conserved (Furukawa *et al*., 2005; Miskei *et al*., 2009).

Clearly, the TFs HogA and AtfA play key roles in regulating stress responses in *A. nidulans*, but these responses are fine-tuned by other global regulatory TFs. These include the carbon catabolite repressor, CreA; the nitrogen metabolite repressor, AreA; and the regulator of the cross pathway control of amino acid biosynthesis, CpcA (Dowzer and Kelly, 1991; Wilson and Arst, 1998; Hoffmann *et al*., 2001). Regulation at the post-transcriptional level, which has been documented for both CreA and AreA, may contribute significantly to mediation of stress responses by TFs such as HogA and AtfA (Small *et al*., 1999; Strauss *et al*., 1999). Interestingly, CpcA is the orthologue of the *S. cerevisiae* TF *GCN4*, which plays a central role in adaptation to diverse stress responses and induces expression of up to 10% of the genome in response to these stresses (Hinnebusch and Natarajan, 2002).

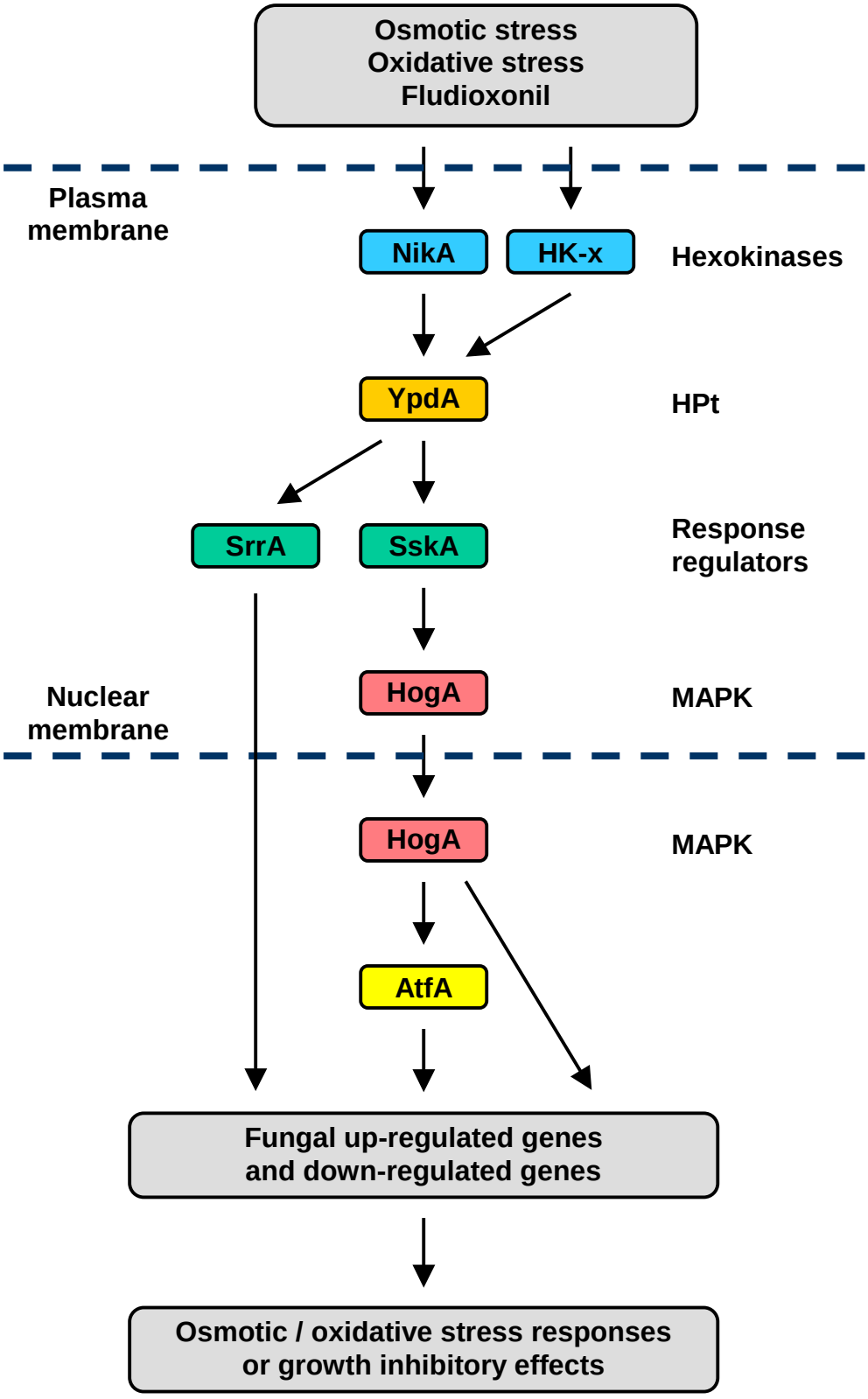


Figure 1.3. A model for His–Asp phosphorelay signalling and the HogA–AtfA MAPK cascade in *A. nidulans*. Adapted from Hagiwara *et al*. (2009) and Lara-Rojas *et al*. (2011). The hexokinase (HK) NikA serves as a sensor that responds to osmotic stress, oxidative stess and the fungicide fludioxonil. NikA transmits the signal *via* YpdA (histidine-containing phosphotransfer intermediate; HPt) to SskA and SrrA (response regulators; RR). The signal is then transduced to the MAPK HogA, which localises to the nucleus where it interacts with AtfA to initiate a transcriptional response. The SskA–HogA–unknown factor are required for the growth inhibitory effect of fludioxonil. Details of the pathway involving SrrA have yet to be established.

1.4 Zinc finger transcription factors

Eukaryotic transcription factors (TFs) function by recognition of specific DNA sequences and are characterised on the basis of their DNA-binding domains (reviewed in Laity *et al*., 2001). The *A. nidulans* genome contains genes encoding 475 TFs, which contribute to twelve classes of DNA-binding proteins (Wortman *et al*., 2009). The vast majority of these TFs are zinc finger proteins that fall into three classes: Zn_2Cys_6 (330), C_2H_2+Zn_2Cys_6 (12) and C_2H_2 (60). Zinc finger proteins are a large group of eukaryotic proteins with diverse roles in cellular processes that include transcription and translation, DNA replication and repair, metabolism, cell signalling and proliferation, and apoptosis (Iuchi, 2001).

The C_2H_2 zinc finger is one of the most common DNA-binding motifs in eukaryotes (reviewed in Iuchi, 2001; Larabee *et al*., 2005). A single C_2H_2 zinc finger is a small peptide domain composed of a β-hairpin and an α-helix, which occur at the N- and C-terminal regions of the finger, respectively. This secondary structure is stabilised by a zinc ion bound between the paired cysteines (within the β-hairpin) and histidines (within the α-helix). Additionally, the zinc finger structure is stabilised by interactions between the conserved hydrophobic residues. The C_2H_2 zinc finger motif is shown in Figure 1.4 (reviewed in Knight and Shimeld, 2001). The zinc fingers wrap around the outside of the DNA helix following a right-handed helical path. The α-helix of each finger contacts the DNA and multiple contacts are made with nucleotide bases in the major groove. Frequently, the amino acid residues within the α-helix interacting with the DNA are at positions -1, +2, +3 and +6 relative to the first amino acid residue of the α-helix. Strong preferences have been observed for the binding of particular amino acids to particular bases, but at present these observations are insufficient to define a set of coding rules. Zinc fingers are also known to function in RNA (Brown, 2005) and protein (Gamsjaeger *et al*., 2007) recognition, in addition to their function as sequence-specific DNA-binding motifs, which is a reflection of their versatility in structure and function. Proteins with several fingers may bind to DNA, RNA or proteins, and different fingers might mediate different types of binding activities (Iuchi, 2001).

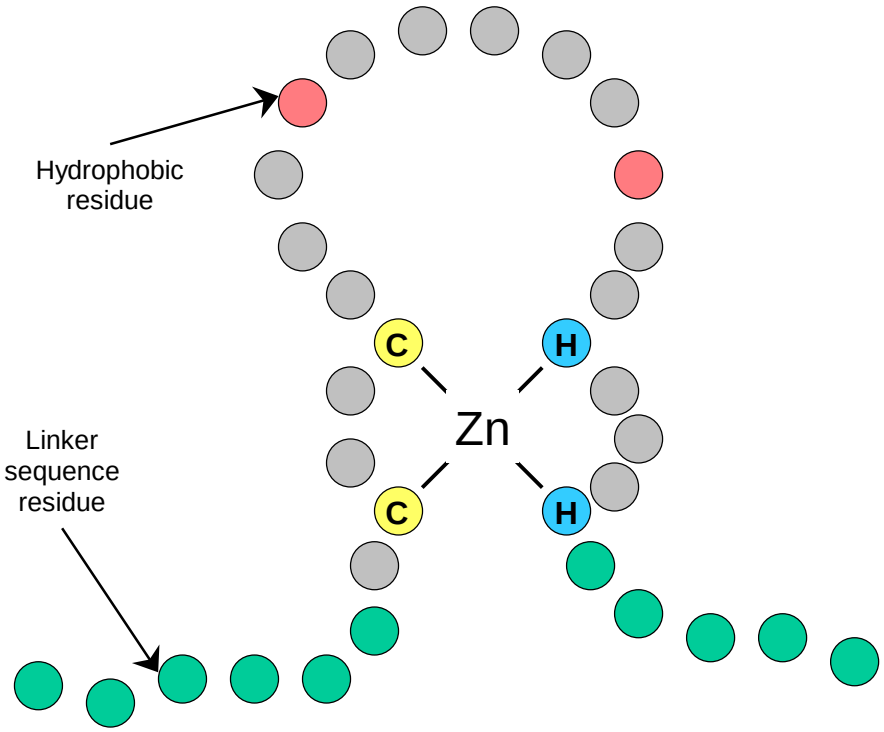


Figure 1.4. The C_2H_2 zinc finger motif. Adapted from Knight and Shimeld (2001). The paired cysteines (C) and histidines (H) that bind the zinc ion are shown in yellow and blue, respectively. The linker sequence, shown in green, joins adjacent fingers. The grey residues are not structurally important and include those responsible for contacting DNA during sequence-specific binding. The precise number of grey residues may vary between the cysteines, the histidines and on the loop. The two large hydrophobic residues, which are structurally important, are shown in red.

1.4.1 The StzA transcription factor in *A. nidulans*

The *sltA1* mutation in *A. nidulans* was initially identified as conferring Na^+ and K^+ sensitivity on the strain GO281 (Spathas, 1978). *A. nidulans sltA1* strains grew poorly on 0.5 M NaCl / KCl, whereas wild-type strains grew into mature colonies in up to 4 M NaCl. The *sltA1* mutation is now known to be pleiotropic by conferring arginase sensitivity, altered arginase activity and a diminished ability to repair DNA damage caused by UV light and the alkylating agent *N*-Methyl-*N*'-Nitro-*N*-Nitrosoguanidine (MNNG), in addition to conferring a reduced ability to induce polyol formation when challenged by salt stress (Clement *et al*., 1996; O'Neil *et al.*, 2002). This research group identified a genomic clone that complemented each of the phenotypic effects conferred by the *sltA1* mutation. Subsequent DNA sequencing and bioinformatic analysis revealed that the clone encoded a TF comprising 698 amino acids with a C_2H_2 motif, which was named StzA (salt tolerant zinc finger protein A). It was determined that the *sltA1* mutation in *A. nidulans* GO281 lacks a (completely) functional *stzA* gene due to the presence of a point mutation at nucleotide 2962 (G>A), which introduces a premature STOP codon at amino acid 502 (W>X). The resultant mutant protein showed loss of the C-terminal region immediately following the zinc finger region, which includes a domain rich in proline residues that is thought to have a role in transcriptional activation. This region also contains one TPXX and two SPXX motifs, which are frequently found in DNA-binding proteins, including CreA, which mediates carbon catabolite repression in filamentous fungi. StzA binds to the sequence 5'-AGGCA in the promoters of target genes (Spielvogel *et al*., 2008). It is noteworthy that spores of *A. nidulans sltA1* strains demonstrate abnormal ability to swell and germinate at acidic pH values when compared to wild-type strains (O'Mahony *et al*., 2002). This behaviour resembles the ability of *A. fumigatus* spores to germinate in the acidic environment of engulfing phagosomes and therefore the *stzA* gene may be involved the perception of favourable conditions to determine when a spore breaks dormancy to colonise a habitat (Waslynka and Moore, 2003). Clearly, StzA is an important TF that controls responses to range of environmental stresses. It is an obvious contender as a key component of a novel mechanism regulating the osmotic response proposed by Furukawa *et al*. (2005).

1.4.2 Proposed allelism of *stzA*/*sltA1* with *agaA*

It was proposed that *sltA1* is an allele of the arginase gene (*agaA*) that produces a salt-sensitive phenotype (Clement *et al*., 1996). *sltA1* strains exhibited substantially reduced sporulation and growth when provided with L-arginine concentrations at or above 10 mM as a sole nitrogen source. Under these conditions, production of a diffusible red pigment, indicative of a stress response, was also evident for *sltA1* strains. This arginine-sensitive response (but not the salt-sensitive response) could be relieved by the addition of L-ornithine, the first product in the catabolism of L-arginine by arginase, to the growth medium.

It has been postulated that *sltA1*/*agaA* is involved in the salt stress response in *A. nidulans* by acting as a modulator of Na^+/H^+ antiporter activity at the plasma membrane. It was shown that Na^+/H^+ exchanger activity was inhibited in response to increasing intracellular arginine concentrations in the presence amiloride, a known inhibitor of the Na^+/H^+ exchanger with remarkable similarity in chemical structure to arginine. It was therefore suggested that arginine is capable of occupying the binding site on the Na^+/H^+ exchanger to modulate its activity (Attwell, 1995; Woodcock, 1997). Hence, the presence of a high extracellular NaCl concentration ultimately leads to an increased intracellular Na^+ concentration: a strategy that allows the cell to equilibrate with hyperosmotic conditions to prevent water loss. However, to reduce Na^+ to a level below that which is toxic to metabolic processes, increased arginase activation occurs to remove inhibitory arginine from the binding sites of plasma membrane Na^+ transporters, such as Na^+/H^+ exchangers, to allow Na^+ to be actively pumped out of the cell. Meanwhile, compatible solutes (e.g. glycerol) are synthesised to prevent water loss in the presence of reduced intracellular Na^+. Furthermore, the products of arginine metabolism ultimately lead to the production of the compatible solute proline and the polyamines putrescine, spermidine and spermine. Polyamines are essential for the normal growth and development of cells in a wide range of organisms (Ruiz-Herrera, 1994) and are known to be required for both developmental transitions, mycotoxin production and stress tolerance in *Aspergillus* species (Walters *et al*., 1997; Jin *et al*., 2002).

Although it has been postulated that *stzA*/*sltA* is allelic with *agaA* (Clement *et al*., 1996), the cloning of *stzA* challenged this hypothesis by showing that *sltA1* represented a mutation in a TF that controlled not only responses to arginase expression but also responses to several other stresses (Section 1.4.1; O'Neil *et al*., 2002). Furthermore, it has yet to be confirmed that *sltA* = *stzA*. These hypotheses therefore need to be formally demonstrated with *stzA* gene knockouts. As StzA appears to control responses to such a wide

variety of stresses, including high arginine concentrations, osmotic and DNA stresses, it is an important TF for further study (O'Neil *et al*., 2002). This is especially important considering that osmotic stress responses between *A. nidulans* and *S. cerevisae* are likely to be very different, despite the fact that they share components of the HOG pathway, owing to their profoundly different lifestyles: *S. cerevisae* inhabits a high sugar environment and is unicellular, whereas *A. nidulans* inhabits the soil and is multicellular (Hagiwara *et al*., 2009).

1.4.3 The StzA orthologue Ace1 in *Trichoderma reesei*

The *ace1* gene product (Ace1) in *Trichoderma reesei* (*Hypocrea jecorina*) demonstrates similarity (58% overall identity) to the StzA protein in *A. nidulans* (O'Neil *et al*., 2002; Aro *et al*., 2003). Furthermore, Ace1 contains three C_2H_2 zinc finger motifs similar to those found in the StzA protein. The *ace1* gene was isolated *via* a yeast based screening system for factors which bound to the *T. reesei* cellulase (*cbh1*) promoter and *in vitro* analysis identified 5'-AGGCA as a consensus recognition site (Saloheimo *et al*., 2000). This is identical to the StzA binding site motif identified for *A. nidulans* StzA (Spielvogel *et al*., 2008). Deletion of the *ace1* gene led to reduced growth on cellulose and implied its role in cellulase regulation. Aro *et al*. (2003) concluded that Ace1 was a repressor of cellulase and xylanase expression, and strains carrying the *ace1* deletion demonstrated impaired sorbitol utilisation. It was concluded that *ace1* had a role in the regulation of some other, as yet, undefined genes. Hence two apparent orthologues, StzA (*A. nidulans*) and Ace1 (*T. reesei*), seem to control a range of distinct phenotypes encompassing abiotic stress response and cellulose/xylan utilisation. The Ace1 protein shows 48% identity and 75% overall similarity with an orthologue in *N. crassa*, and the region containing the zinc fingers and putative nuclear localisation signal is nearly identical among all three Ace1/StzA orthologues (Aro *et al*., 2003).

1.5 Gene targeting in filamentous fungi

An increased understanding of fundamental genetic processes in model organisms has resulted in the development of new strategies for homologous recombination (HR) by DNA-mediated transformations (reviewed in Krappmann, 2007; Larrondo *et al*., 2009; Wortman *et al*., 2009; Kück and Hoff, 2010; Haas *et al*., 2011). These transformations result in either a gene knock-out or gene knock-in. With knock-out transformations, the target gene is substituted with a marker gene (e.g. antibiotic resistance or auxotrophic marker genes). With knock-in transformations, the target gene is substituted with a foreign or mutated gene.

Until recently, DNA-mediated transformations in filamentous fungi have been difficult because self-replicating vectors are rare in filamentous fungi and the transferred DNA is often ectopically integrated into the genomic DNA (Kück and Hoff, 2010). Furthermore, DNA integration in filamentous fungi mainly occurs by non-homologous end-joining (NHEJ), and the site-specific recombination necessary for effective analysis of gene function occurs at very low frequencies (less than 1%). This meant that gene disruption experiments involved the screening of hundreds of transgenic strains to identify the desired disruption strains, a process that was laborious and time-consuming.

Such technical difficulties have now been largely overcome by the introduction of Polymerase Chain Reaction (PCR)-based gene targeting using split-marker technology, which was initially developed for *S. cerevisiae* (Fairhead *et al*., 1996) but has since been successfully applied to *A. nidulans* (Nielsen *et al*., 2006; Nayak *et al*., 2006; Szewczyk *et al*., 2006) and other filamentous fungi (Krappmann, 2007; Larrondo *et al*., 2009). This split-marker technology involves fusion PCR, whereby marker gene fragments are fused with DNA fragments identical to those that flank the gene to be deleted. Two bipartite gene-targeting substrates are generated, which combine homologously with the target DNA region *in vivo* during transformation. During this process, the marker gene replaces the targeted gene.

The sequencing of the genomes of filamentous fungi has enabled gene sequences to be identified and their flanking regions to be amplified by PCR using primers based on genomic DNA sequences (Wortman *et al*., 2009). Hence, although the genes themselves are not amplified, their flanking sequences must be known. Since more than 100 fungal genome sequences have now been published, there is great potential for targeted gene manipulations to be performed to increase our understanding of gene functions in this group

of organisms containing species of medical and industrial importance (Kück and Hoff, 2010; Haas *et al*., 2011). It is noteworthy that a major project is underway to use gene-targeting technology to systematically delete each of the 10,000 predicted genes in the *N. crassa* genome (Dunlap *et al*., 2007).

1.6 Aims

The main overall aim was to characterise the function of *A. nidulans* StzA by construction and phenotypic analyses of *stzA* deletion strains. Additionally, identification and bioinformatic analyses of StzA orthologues enabled sequence comparisons. Promoter analysis enabled potential interactions of StzA with other TFs to be deduced.

Chapter 2

Materials and Methods

2.1 Origins of fungal and bacterial strains and plasmids

Fungal and bacterial strains and plasmids used during the project were maintained at the University of Wolverhampton. *A. nidulans* strains were originally supplied by the University of Liverpool (L19, L20 and G191), Dr. A. J. Clutterbuck from the University of Glasgow (GO281), D. J. Clement (WV101 and WV202) and the Fungal Genetics Stock Centre (A1149). *T. reesei* strains QM9414 and VTT-D-061244 (*ace1* deletion) were provided by Nino Aro and Merja Pentillä from VTT Biotechnology, Finland. Plasmid pDEL1 was supplied by Michael L. Nielsen and Uffe H. Mortensen from the Technical University of Denmark and plasmid pDJB3 by Prof G. Turner from the University of Sheffield. Strains and plasmids used, along with their genotypes, are listed in Appendix 1.

2.2 Maintenance of fungal and bacterial strains

A. nidulans and *T. reesei* strains were cultured and maintained on Petri plates with MEA (2% malt extract and 1.5% agar). Hydrochloric acid (1 M) or sodium hydroxide (1 M) was used to adjust the pH to 6.5. *A. nidulans* strains carrying the *pyrG89* mutation were grown on MEA supplemented with 10 mM uridine and 10 mM uracil. All strains were incubated at 37 ºC for 3 to 4 days to generate stock plates. Glycerol stocks of spore suspensions (SDW inoculated with ~1 x 10^8 spores ml^{-1} with the addition of 50% glycerol) were frozen at –70 ºC for long-term storage (Clutterbuck, 1974).

Escherichia coli strain JM109 was cultured on Luria Bertini Agar (LBA; Appendix 2). Glycerol stocks of cell suspensions were created from aliquots of LB medium inoculated with ~1 x 10^9 cells ml^{-1} and incubated at 37 ºC (200 rpm) prior to the addition of 50% glycerol. These cell suspensions were then frozen at –70 ºC for long-term maintenance (Sambrook and Russell, 2001).

2.3 Growth of fungal strains on minimal media

Growth of *A. nidulans* strains on different carbon sources was assessed on Petri plates with minimal medium (MM) containing 2 mM $MgSO_4.7H_2O$, 7 mM KCl, 10 mM KH_2PO_4, 10 mM $C_4H_{12}N_2O_6$ (ammonium tartrate used as a nitrogen source), 1 ml trace elements solution and 1 ml vitamin solution (Appendix 2; Cove, 1976). This MM was supplemented with 1% (w/v) of a particular carbon source. Glucose was used as the sole carbon source for experiments testing sensitivity to stresses (e.g. cations and antifungal compounds). *A. nidulans pyrG89* strains were grown on MM supplemented with 10 mM uridine and 10 mM uracil. Strains carrying *pabaA1* or *pyroA4* mutations were grown on MM containing final concentrations of 5 µM para-amino benzoic acid or 0.25 µM pyridoxine. The pH of media was adjusted to pH 6.5 using 1 M hydrochloric acid or 1 M sodium hydroxide. Agar (1.5%) was added before the solution was autoclaved, prior to pouring. Plates were aseptically inoculated from stock plates of fungal spores by spot transfer. Two plates were spot inoculated three times so that there were six replicates for each experiment. Plates were incubated at 37 ºC for either 48 or 72 hours. *T. reesei* strains were grown on amended MM or MEA (2%) at 30 ºC for 72 hours. Photographs of the plates were taken (by Dr. M. Inman) to provide permanent records of the morphological states of the fungus, enabling growth of strains under various conditions to be compared with growth on glucose.

2.4 DNA damage sensitivity tests

Sensitivity of *A. nidulans* and *T. reesei* strains to UV was assessed by the method of Hooley *et al*. (1988). Colonies were grown by spot inoculating MEA plates and incubating them at 37 ºC (*A. nidulans* strains) or 30 ºC (*T. reesei* strains) for 24 hours. These colonies were irradiated with UV light for 6 minutes and promptly returned to the incubator for a further 24 hours. Photoreactivation tests were performed by placing plates with 24 hour-old colonies in conditions of light for 2 hours immediately following UV treatment (6 mins), prior to 24 hours of further incubation. Colonies subjected to two hours of dark treatment were used as a control. Sensitivity of fungal strains to the UV mimic 4-nitroquinoline-1-oxide (4NQO) and the alkylating agent *N*-Methyl-*N*'-Nitro-*N*-Nitrosoguanidine (MNNG) was performed by spot inoculation of MEA agar plates containing these two DNA-damaging agents (Hooley *et al*., 1998). Stock solutions of MNNG (12.5 mg ml^{-1}) and 4NQO (3 mg ml^{-1}) were made up in ethanol at –20 ºC and kept for up to 2 days. Aliquots of the stock

solutions were added to molten MEA. Plates were incubated at 37 ºC for 48 hours and 30 ºC for 72 hours for *A. nidulans* and *T. reesei* strains, respectively.

2.5 Sexual crosses between *A. nidulans* strains

Sexual crosses of *A. nidulans* strains were based on the method of Pontecorvo *et al*. (1953) and were set up between strains that were of differing colours and complementary auxotrophies. Tiger plates were set up in which the two strains to be crossed were introduced onto one MEA plate by alternately spot inoculating the strains along the equator of the plate. Thus, following incubation of such plates for 2 days at 37 ºC, a line of overlapping colonies, which alternated with respect to strain, was clearly visible due to the colour differences between the strains. At the overlapping regions, small 2–3 mm agar plugs were excised and placed onto MM plates that lacked a differing essential vitamin (para-amino benzoic acid, paba; or pyridoxine, pyro). Approximately 10–15 agar plugs were dispersed over each plate and then incubated for several days at 37 ºC. During this incubation period, heterokaryons formed and were visible as dense, fast growing sectors, which initially appeared brown before the emergence of both parental colours. At that point, the plates were sealed with parafilm, before being returned to the incubator for 2–3 weeks at 37 ºC. During this time, oxygen and nutrient starvation forced diploids/meiosis, resulting in the formation of black, shiny, mature cleistothecia, which contained the ascospores. Cleistothecia were covered with a layer of light yellow Hülle cells and parental conidia. These cells were removed by rolling cleistothecia on plates containing 3% distilled water agar (DWA). When clean, each cleistothecium was transferred to an Eppendorf tube containing 500 µl sterile distilled water (SDW), crushed and vortexed to release the ascospores. These ascospores were tested for evidence of outcrossing by transferring different volumes to fully supplemented MEA plates. Cleistothecia that were true hybrids (not selfed or twinned) could be identified by looking for the expected spore colour ratios. An appropriate dilution of ascospores from true hybrids was used to select single colonies, which were used to form master plates, each containing 10 colonies. As neighbouring spores might produce the same colour conidia, even single-coloured colonies were picked near an edge. Master plates were incubated at 37 ºC for 2 days to obtain discrete colonies. Any plates containing overlapping colonies were discarded. Growth morphologies of progeny were investigated by transfer to amended MM providing different sole carbon sources or exerting various stresses. Typically, each treatment was carried out with combinations of the presence or absence of either or both essential nutrients (i.e. the presence or absence of paba and/or

pyro) to confirm that the crosses had been successful (i.e. segregation of these markers in a 1:1 ratio) and to identify any inconsistencies. The colonies were always viewed and photographed following 48 hours of incubation at 37 ºC.

2.6 Sterigmatocystin extraction and analysis by Thin Layer Chromotography

Sterigmatocystin production by *A. nidulans* L20 and GO281 was examined by Thin Layer Chromotography (TLC) analysis using the method described by Stack and Rodricks (1971). *A. nidulans* cultures were set up by spot inoculation of plates containing 2% MEA and 1.5% agar. Growth of both strains was similar after 72 hours of incubation at 37 ºC. Haemocytometer counts showed that these plates contained ~5 x 10^7 spores. SDW (10 ml) was added to each spore plate, before gently scraping the spores into the water and transferring the resulting spore suspensions into test tubes. The spore suspensions were mixed thoroughly by vortexing, before 1 ml of each suspension was transferred to 2 ml Eppendorf tubes. Chloroform (1 ml) was added and mixed by vortexing. Following centrifugation for 5 minutes at 1000 rpm, the separated organic phase was transferred to an Eppendorf tube. The samples were dried by allowing the chloroform to evaporate in a fume cupboard overnight. The following day, the samples were resuspended in 100 µl of chloroform. Aliquots (5 µl and 25 µl) of each extract along with sterigmatocystin standards (5–1000 ng) dissolved in benzene were spotted onto a silica gel TLC plate and separated in toluene – ethyl acetate – formic acid (40 : 60 : 0.5). The TLC plate was sprayed with aluminium chloride to enhance sterigmatocystin fluorescence upon exposure to longwave (365 nm) UV light.

2.7 Preparation of competent bacterial cells

Competent bacterial cells were prepared and transformed using an adaptation of the method of Hanahan (1985). The *E. coli* strain JM109 was chosen for these procedures because its *recA*$^{-}$ genotype prevents undesirable recombination between the plasmid and host chromosomal DNA. Additionally, the endonuclease A^{-} mutation (*endA*$^{-}$) leads to an improved yield and quality of isolated plasmid DNA, and the *hsdR* mutation prevents the cleavage of cloned DNA by the EcoK endonuclease system (Inoue *et al*., 1990). A streak plate containing LBA (Appendix 2) was prepared using *E. coli* JM109 and incubated overnight at 37 ºC. The following day, a single colony from the streak plate was used to inoculate an LB starter culture (25 ml in a 250 ml conical flask), which was incubated at 37 ºC on an orbital shaker (225 rpm) for approximately 16 hours. The following day, this entire starter culture was transferred to a 1-litre conical flask containing 225 ml of LB broth with 20 mM magnesium sulphate. The flask was incubated at 37 ºC on an orbital shaker (225 rpm) until the A_{600} of the culture was between 0.4–0.6 (~3 hours). The cells were then pelleted by centrifugation at 4,500 g for 5 minutes (4 ºC). Henceforth, the competent cells were treated carefully due to their high sensitivity to handling and elevated temperatures. The supernatant was poured off and the cell pellet was gently resuspended in 50 ml of ice cold TFB1 (Appendix 2) using a glass rod. The resuspended cells were centrifuged again at 4,500 g for 5 minutes (4 ºC). Again, the supernatant was poured off and the cell pellet was resuspended in 10 ml of ice cold TFB2. The cell suspension was incubated on ice for 30 minutes. Aliquots (300 µl) of the cell suspension were subsequently transferred to Eppendorf tubes and flash frozen by immersion in liquid nitrogen.

2.8 Plasmid transformation of competent bacterial cells

Competent cells (250 µl) were added to Eppendorfs containing 5 µl of the plasmid pDJB3 or pDEL1 (10 ng µl^{-1}). Plasmid DNA was mixed with the cells by gently drawing the suspension through a pipette tip prior to incubation on ice for 30 minutes. The cells were then heat shocked at 42 ºC for 3 minutes. One ml of LB medium was added to the cells, which were incubated at 37 ºC for 45 minutes to allow expression of the ampicillin resistance genes, which both pDJB3 and pDEL1 contain. Transformed cells were then spread over LB agar plates with 100 µg ml^{-1} ampicillin. When the plates were dry, they were inverted and incubated overnight at 37 ºC. Since JM109 cells have no natural resistance to ampicillin, only competent cells that took

up the plasmids conferring ampicillin resistance formed colonies on the ampicillin plates that were visible the following day. Identical transformation procedures were performed using competent cells mixed with plasmid pUC19 (50 ng) as a positive control and competent cells with no added plasmid DNA as a negative control. Aliquots of each transformation mixture were also plated on LB plates containing no ampicillin.

The plasmids pDJB3 and pDEL1 were amplified by transferring single transformed JM109 colonies to 10 ml of LB medium (with 100 µg ml^{-1} ampicillin) in a 50 ml Universal tube and grown overnight at 37 ºC (225 rpm). Aliquots (0.4 ml) of the cell suspension were transferred to Eppendorfs and 0.6 ml of 75% glycerol was added prior to storage at –70 ºC.

2.9 Plasmid DNA isolation and purification

Plasmid DNA was isolated and purified using a Qiagen Plasmid Maxi Kit, which involves an adaptation of the alkaline lysis method described by Sambrook and Russell (2001). Components of all buffers used are listed in Appendix 2. Streak plates were produced containing *E. coli* JM109 cells transformed with either the plasmid pDJB3 or pDEL1 from glycerol stocks. A single colony from a freshly prepared streaked plate was used to inoculate a 50 ml conical flask containing a 5 ml starter culture of LB medium with 100 µg ml^{-1} ampicillin. This starter culture was incubated at 37 ºC for 8 hours in an orbital shaker with vigorous shaking (300 rpm). The starter culture was subsequently diluted 1/500 into selective LB medium. Thus, 0.5 ml of the starter culture was used to inoculate 250 ml of LB medium with 100 µg ml^{-1} ampicillin in a 1-litre conical flask. This culture was incubated at 37 ºC for 12–16 hours with vigorous shaking (300 rpm).

The following day, the bacterial cells were harvested by centrifugation at 6000 x *g* for 15 minutes at 4 ºC. The cell pellet was then resuspended in 10 ml of Buffer P1. This was followed by the addition of 10 ml of Buffer P2, which was mixed vigorously by rapidly inverting the tube 4–6 times and incubated for 5 minutes at room temperature. This mixing procedure was repeated following the addition of 10 ml of chilled Buffer P3, and the tube was incubated on ice for 20 minutes. Centrifugation was then carried out at 20,000 x *g* for 15 minutes at 4 ºC. The supernatant, containing the plasmid DNA, was removed promptly, transferred to another tube and centrifuged again at 20,000 x *g* for 15 minutes at 4 ºC. The supernatant was again removed promptly. A Qiagen tip-500 was equilibrated by applying 10 ml of Buffer QBT and the column was

allowed to empty by gravity flow. The supernatant was then poured into the Qiagen tip and the column was again allowed to empty by gravity flow. The Qiagen tip was washed with 2 x 30 ml of Buffer QC. Buffer QF (15 ml) was used to elute the DNA. The DNA was precipitated by adding 10.5 ml (0.7 volumes) of isopropanol and centrifuged immediately at 15,000 x *g* for 30 minutes at 4 ºC. The supernatant was carefully decanted and the DNA pellet was washed in 5 ml of 70% ethanol prior to centrifugation at 15,000 x *g* for 10 minutes. The supernatant was again carefully decanted and the pellet air-dried for 10 minutes before the DNA was resuspended in 100 µl of 10 mM Tris-HCl, pH 8.5.

2.10 Isolation of *A. nidulans* total DNA

Isolation of total DNA from *A. nidulans* was performed using an adaptation of the method of Raeder and Broda (1985). All centrifugation steps were carried out at 17,900 x *g* (13,000 rpm) using a table top microcentrifuge at room temperature. Overnight cultures of *A. nidulans* were prepared by aseptically scraping fresh 4-day-old spores (~1 x 10^8) from an MEA Petri plate into a conical flask (500 ml) containing 150 ml of 2% ME broth. The cultures were incubated at 37 ºC on orbital shakers (200 rpm) overnight (~16 hours). The following day, the mycelial biomass from each culture was extracted and dried by filtering the cultures through sterile muslin followed by squeezing of the biomass to remove excess liquid. The biomass was added to a mortar containing liquid nitrogen and ground into a fine powder using a pestle. Extraction buffer (2 ml), phenol:chloroform:isoamyl alcohol (25:24:1; 2 ml) and sodium dodecyl sulphate stock solution (2 ml) were added to the mortar and mixed (Appendix 2). Aliquots of the mixture were added to sterile Eppendorfs, which were centrifuged at 13,000 rpm for 20 minutes. The upper aqueous phase from each tube was transferred to Eppendorfs containing 10 µl of RNAseA (final concentration of 100 µg ml^{-1}) and then incubated at 37 ºC for 30 minutes. Following incubation, RNAseA was removed by the addition of an equal volume of phenol:chloroform:isoamyl alcohol to each Eppendorf, prior to centrifugation at 13,000 rpm for 5 minutes. The upper aqueous phases were then transferred to fresh Eppendorfs. The phenol/chloroform extraction process was repeated 2–3 times until the upper aqueous phase was completely clear. Once clear, the upper phase was transferred to fresh Eppendorfs. Following the addition of 0.6 volumes of isopropanol, the DNA precipitated visibly when the Eppendorfs were inverted. Centrifugation was carried out at 13,000 rpm for 5 minutes and the liquid was poured off the DNA pellets. The Eppendorfs were spun again at 13,000 rpm for 5 minutes and any remaining liquid was poured off. Ethanol (1 ml; 70%) was added to each

Eppendorf before centrifugation was carried out at 13,000 rpm for 2 minutes. The ethanol was subsequently poured off and the pellet dried in an air safety cabinet for 30–50 minutes. When dry, 200 µl of 10 mM Tris-HCl pH 8.0, 1 mM EDTA (or SDW) was added to each Eppendorf. The pellets were resuspended by gently tapping the tubes. Once resuspended, the *A. nidulans* total DNA samples were stored at –20 ºC until required.

2.11 Determination of DNA concentration and purity

Concentrations and purities of DNA samples were estimated initially by UV spectrophotometry using the method of Sambrook and Russell (2001). Ten µl of the DNA sample was added to 1990 µl of SDW (200x dilution). This was added to a quartz cuvette and the absorbance was read at 260 nm (against an SDW blank). An optical density of 1 corresponds to a DNA concentration of 50 µg ml^{-1}. DNA purity was estimated by comparing the A_{260} and A_{280} absorbance readings. A 260/280 ratio of 1.8 is indicative of pure DNA. Concentrations of genomic DNA, plasmid DNA or PCR-amplified DNA were more accurately determined by running serial dilutions of DNA samples on an ethidium bromide-stained gel (Sambrook and Russell, 2001). A UV transilluminator was then used to visualise and compare the thicknesses and intensities of the DNA bands with those of known concentrations (i.e. bands in 5 µl of DNA Hyperladder I (720 ng) from Bioline or 1 µg of Lambda DNA digested fully with *Hin*dIII). The purity of genomic DNA samples was also determined visually by observing the extent of genomic DNA restriction following overnight digests with a single restriction enzyme.

2.12 DNA restriction digests

Genomic DNA digests were used to assess the purity of genomic DNA samples prior to their use in PCR. Restriction digests were also used to verify that desired DNA fragments had been amplified by PCR. All restriction enzyme digests were performed using the buffers recommended by the manufacturer. Typical reaction conditions for genomic DNA and PCR-generated DNA fragment digests, using a single restriction enzyme, are shown below:

Typical genomic DNA digest (20 µl)

X µl SDW

X µl genomic DNA sample (~500 ng)

2 µl 10x restriction buffer

2 µl restriction enzyme (10 units)

Incubated at 37 ºC overnight

Typical PCR-generated DNA fragment digest (20 µl)

X µl SDW

X µl DNA sample (200–300 ng)

2 µl 10x restriction buffer

2 µl restriction enzyme (10 units)

Incubated at 37 ºC for 4–6 hours

2.13 Agarose gel electrophoresis

Agarose gels (0.8%) were cast using 50 or 100 ml of Tris-Borate EDTA (TBE) buffer with the addition of ethidium bromide (25 or 50 µl of 1 mg ml^{-1}) to a final concentration of 0.5 µg ml^{-1} (Appendix 2). Gel loading buffer (6x) was added to DNA samples prior to electrophoresis, whereby DNA was fractionated at 1–5 V cm^{-1} (Sambrook and Russell, 2001).

2.14 DNA amplification using the Polymerase Chain Reaction

Specific DNA sequences were amplified from DNA templates using the Polymerase Chain Reaction (PCR; Mullis *et al*., 1986). Reaction and thermocycler conditions used for the synthesis of gene-targeting substrates for the deletion of the *stzA* gene are detailed specifically in Chapter 4.

PCR primers were designed using the Primer3 program (http://frodo.wi.mit.edu/primer3/). When designing primers, considerations were given to the following criteria (Sambrook and Russell, 2001):

- Primers should be 18–30 nucleotides in length.
- Primers should be unique.
- Pairs of primers should have similar annealing temperatures (within 5 ºC of each other, and therefore similar in G+C content). The melting temperatures (T_m) for oligonucleotides could be estimated using the Wallace rule: T_m (ºC) = 2 (A+T) + 4 (G+C). The annealing temperature was initially set approximately 5 ºC below the lowest T_m of the primer pair.
- Primers should have a minimal degree of self-complementarity (to avoid formation of secondary structures).
- Pairs of primers should have minimal complementarity to each other (to avoid formation of primer dimers).
- Sequences with long runs of a single nucleotide (i.e. more than 3 or 4) should be avoided.
- The last two bases at the 3' end of the primer should ideally be either G or C (allowing 3 hydrogen bonds to be formed for stronger annealing than the 2 hydrogen bonds formed between A and T).

2.15 Purification of PCR-amplified DNA fragments

DNA fragments amplified by PCR were purified with a Qiagen QIAquick Gel Extraction Kit. Components of buffers used are listed in Appendix 2. All centrifugation steps were carried out at 17,900 x *g* (13,000 rpm) using a table top microcentrifuge at room temperature. Multiple gel slices containing the desired DNA fragment were excised from an agarose gel using a scalpel. The gel slices were weighed, added to a 1.5 ml Eppendorf, and 3 volumes of Buffer QG were added to 1 volume of gel (i.e. 300 µl of Buffer QG was added

to each 100 mg of gel). Eppendorfs were incubated at 50 ºC until the gel slices had completely dissolved (approximately 5–10 minutes). One gel volume of isopropanol was added to each sample (i.e. 100 µl of isopropanol for each 100 mg of gel) and mixed. The sample was added to a QIAquick spin column placed inside a 2 ml collection tube and centrifuged for 1 minute. The flow-through was then discarded and the QIAquick column was placed inside the same 2 ml collection tube. Buffer QG (0.5 ml) was added to the column, which was centrifuged for 1 minute to remove all traces of agarose. Buffer PE (0.75 ml) was added to the column, which was allowed to stand for 5 minutes before centrifugation for 1 minute. The flow-through was discarded and the column was centrifuged for an additional minute to remove any residual Buffer PE (containing ethanol). The column was then placed in a clean collection tube and 50 µl of SDW was placed on the centre of the column membrane. The column was allowed to stand for 5 minutes, and then centrifuged for 1 minute to elute the DNA. The purified DNA was stored at –20 ºC.

2.16 Preparation of *A. nidulans* protoplasts

A. nidulans protoplasts were prepared and transformed using modifications to the method of Tilburn *et al*. (1983). Approximately 1 x 10^8 *A. nidulans* spores of strain G191 or A1149 were used to inoculate a 500 ml conical flask containing 200 ml of liquid MM supplemented with vitamins and 10 mM of uridine and uracil. Following incubation of the cultures at 30 ºC for 16 hours in an orbital shaker (120 rpm), the biomass was harvested in sterile muslin and transferred to a 100 ml conical flask containing 15 ml 0.6 M KCl buffer containing 20 mg ml^{-1} Lysing enzyme from Sigma (L1412). Following gentle shaking at 30 ºC for 2 hours, protoplasts were harvested and filtered through a sintered glass funnel to remove mycelial debris. The filtrate, containing the protoplasts, was spun down at 2,500 g in a swing out rotor for 5 minutes. Protoplasts were visible as dark spheres when viewed under a phase contrast microscope.

Cell wall removal from spores to yield protoplasts could be verified by observing lysis of protoplasts following the addition of SDW to protoplast samples. The protoplasts were washed twice with 0.6 M KCl (15 ml) and resuspended by shaking gently. Approximately ~5 x 10^7 protoplasts were then washed once in 0.6 M KCl, 50 mM $CaCl_2$ (15 ml) and resuspended in 250 µl of the same solution. Great care was taken to avoid rupturing protoplasts at all stages during the stages of washing/purification and transformation. The ends of pipette tips were cut to minimise rupturing.

2.17 Transformation of *A. nidulans* protoplasts

Plasmid pDJB3, which contains the *pyr4* gene from *N. crassa*, was used to transform the *A. nidulans pyrG89* strains A1149 and G191. Approximately 1 µg of the plasmid (in up to 10 µl of TE buffer) was added to a sterile Bijoux and mixed with 100 µl of the protoplast suspension (~2 x 10^7 protoplasts). Twenty five µl of PEG solution (25% PEG 6000, 0.6 M KCl, 50 mM $CaCl_2$, 10 mM Tris, pH 7.5) was added and mixed instantly. The protoplasts were incubated on ice for 20 minutes. PEG solution (1 ml) was then added and rapidly but gently mixed and then left at room temperature for 30 minutes. Two ml of 0.6 M KCl, 50 mM $CaCl_2$ solution was then added and mixed by gently inverting the Bijoux. Aliquots of the transformation mixture (typically 250 µl) were added to aliqouts of regeneration medium overlay (5 ml; Appendix 2) at ~45 ºC, mixed and gently poured into Petri plates containing a base layer of regeneration medium. Alternatively, aliquots were plated directly onto regeneration medium. Plates were initially incubated at 30 ºC overnight and then at 37 ºC for 6 days, when the majority of transformants were visible due to pyrimidine auxotrophy complementation. Stable transformants used as subsequent control strains (AP1, AP2, GP1 and GP2) were purified by streaking out spores onto selective medium to obtain single colonies.

Approximately 200 ng of each bipartite gene-targeting substrate was used to transform 2 x 10^7 protoplasts from *A. nidulans* A1149 and G191 using the method described for transformation with pDJB3. This plasmid was used again as a positive control and an identical procedure was performed with no DNA as a negative control. The stable *stzA* gene deletion transformants obtained (A2, A5, A7, G6 and G25) were purified by streaking out spores onto selective medium to obtain single colonies.

2.18 DNA sequencing

DNA sequencing was performed by the Dundee Sequencing Centre (http://www.dnaseq.co.uk) using the method of direct sequencing. This company sequenced key regions of the PCR assay product (2,437 bp) obtained from the *stzA* gene deletion mutant designated A2 and also designed the primers SA and SB to enable the sequencing reactions to be performed.

2.19 Analysis of plant cell wall degrading enzymes produced by *A. nidulans* strains

2.19a Growth of *A. nidulans* cultures and collection of enzyme samples

One plate of 3-day-old spores (grown on 2% MEA) of either strain A1149 or A2 was harvested in 20 ml of SDW. The spore suspension was centrifuged at 4,500 rpm for 10 minutes. The spore pellet was then washed to remove all traces of MEA and resuspended in 20 ml SDW. Centrifugation was repeated and spores were resuspended in 5 ml SDW. The spores were used to inoculate a 500 ml conical flask containing 100 ml MM (with either 1% of xylan, cellulose or polygalacturonic acid as a sole carbon source) to produce a final spore concentration of 1×10^7 spores ml^{-1}. The cultures were incubated on a rotary shaker at 37 ºC (120 rpm) for 48 hours. Following this incubation period, spores were pelleted (4,500 g for 10 minutes at 4 ºC) to obtain spore-free supernatants containing the desired enzymes. Aliquots were stored in sterile universals at –20 ºC and used to perform the DNS assay (Section 2.19b) and the Bradford protein assay (Section 2.19c).

2.19b Dinitrosalicylic acid assay for reducing sugar determination

The dinitrosalicylic acid (DNS) assay was based on the method of Miller (1959). Aliquots (300 µl) of the appropriate substrate solution (Table 2.1) for the type of reducing sugar being analysed were incubated at 37 ºC for 15 minutes. Either 100 µl or 10 µl (+ 90 µl SDW) of the spore-free supernatant, containing the plant polysaccharide degrading enzymes, were added to the substrate solution. These enzyme reactions (400 µl) were incubated at 37 ºC on a rotary shaker (~150 rpm) for 30 minutes. DNS-lactose solution (800 µl; Appendix 2) was then added to each Eppendorf to stop the reaction. The Eppendorfs were boiled for 15 minutes and cooled by transfer to ice. They were then spun at 10,000 rpm to remove any insoluble substances. Reaction blanks were prepared alongside the test reactions by adding 100 µl of SDW to 300 µl aliquots of the substrate solutions, and then received identical treatment as the test reactions. Absorbances of the test reactions were read at 540 nm against the appropriate reaction blank. Xylose (0.5%), glucose (0.5%) and galacturonic acid (1%) were used as the reducing sugar stock standards and were diluted appropriately with SDW to produce standard curves. Absorbances at 540 nm were again read against the reaction blanks.

Table 2.1. Carbon sources, substrate solutions and standard solutions used for the DNS assay.

C-source used for 2-day culture	DNS assay conditions	
	Substrate solution (1%)	Standard solution
Xylan	Xylan	Xylose (0.5%)
Cellulose	Cellulose	Glucose (0.5%)
Polygalacturonic acid	Polygalacturonic acid	Galacturonic acid (1%)

2.19c Bradford protein assay

The Bradford protein assay was used to assay total protein concentrations in each sample (Bradford, 1976). Bradford reagent (200 µl) was added to either 800 µl, 400 µl or 200 µl of each spore-free supernatant (made up to 800 µl with SDW) to produce 1 ml reaction volumes. SDW (800 µl) was added to Bradford reagent (200 µl) and used as a reaction blank. Test reactions and the blanks were incubated at room temperature for 45 minutes. Absorbances of the test reactions were then read against the reaction blanks at 595 nm. Bovine serum albumin (BSA) was used to produce a standard curve by using appropriate dilutions with 50 mM potassium phosphate buffer, pH 7.8.

2.20 Identification and bioinformatic analysis of StzA orthologues

StzA orthologues were identified by performing blastn and blastp searches at the National Center for Biotechnology Information (NCBI; http://blast.ncbi.nlm.nih.gov/Blast.cgi) and the Broad Institute (http://www.broad.mit.edu/node/568). Search parameters were left on default values. The top matching protein sequence was taken as the orthologue, provided the e-value was less than $3e^{-30}$. DNA and protein sequences were aligned using Advanced TCoffee (http://tcoffee.vital-it.ch/cgi-bin/Tcoffee/tcoffee_cgi/index.cgi; Notredame *et al*., 2000) and ClustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html; Chenna *et al*., 2003). Translations were performed using Javascript DNA translator version 1.1 (http://nbc11.biologie.uni-kl.de/framed/left/menu/auto/right/JDT/). Introns were identified manually by performing DNA and protein sequence comparisons among *stzA* orthologues with consideration to fungal gene/intron structure (Kupfer *et al*., 2004; Rep *et al*., 2006). Secondary structure predictions were obtained for each protein individually using PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/; Mc Guffin *et al*., 2000) and then superimposed onto an Advanced TCoffee

alignment of all 29 StzA proteins. Tertiary structure predictions were attempted for StzA proteins using GenTHREADER (http://bioinf.cs.ucl.ac.uk/psipred/; Jones *et al*., 1999). Conserved domain searches were carried out using the Conserved Domain Database (CDD) at NCBI (http://www.ncbi.nlm.nih.gov/cdd/). Nuclear Localisation Sequences were predicted using the criteria of Dingwall and Laskey (1998). Phylogenetic analysis of StzA proteins was performed using the Phylogeny.fr platform, which used MUSCLE for multiple alignment, Gblocks for automatic alignment curation, PhyML for tree building and TreeDyn for tree drawing (http://www.phylogeny.fr/; Dereeper *et al*., 2008).

2.21 Promoter analysis of *stzA* orthologues and potential StzA target genes

Gene sequences and their corresponding promoters were obtained using accession numbers at NCBI (http://www.ncbi.nlm.nih.gov/) or the Broad Institute (http://www.broadinstitute.org/), and homologous gene sequences were obtained using the BLAST programs available at these websites. Additional information regarding gene annotations could be obtained from The Central *Aspergillus* Data REpository (CADRE; http://www.cadre-genomes.org.uk/). Promoter sequences were defined as sequences 2 kb upstream of the ATG translation start sites. The sequences of experimentally verified TF binding sites were identified from published literature. Javascript DNA translator (http://nbc11.biologie.uni-kl.de/framed/left/menu/auto/right/JDT/) was used to remove any spaces from promoter sequences prior to using the Microsoft Word® "Find" tool to identify potential TF binding sites (forward and reverse complements) on both DNA strands. Alignments of promoter sequences were carried out using Advanced TCoffee (http://tcoffee.vital-it.ch/cgi-bin/Tcoffee/tcoffee_cgi/index.cgi, Notredame *et al*., 2000); ClustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html, Chenna *et al*., 2003); and Mlagan (http://lagan.stanford.edu/lagan_web/index.shtml, Brudno *et al*., 2003) or by manual alignments. A pattern search at the PEDANT database (http://pedant.gsf.de/) at the Munich Information Center for Protein Sequences (MIPS) was performed to search the entire genome of *A. nidulans* for the *REALALE* sequence, CTATCAGGCA, and its reverse complement TGCCTGATAG, which contains overlapping AreA and StzA binding sites (Wilson and Arst, 1998; Spielvogel *et al*., 2008).

Chi-square analysis was used to assess whether promoters of *stzA* genes were significantly enriched for the presence of specific TF binding site motifs when compared to these occurrences within the promoters of a

control group of genes. Chi-square analysis was also used to assess whether the frequency distribution of StzA binding site motifs differed significantly for any groups of *A. nidulans* promoters from genes assigned to functionally related groups (osmotic stress, DNA repair, cation homeostasis, and *REALALE*-containing). This statistical approach reflects the approach used by Coutinho *et al*. (2009) who used chi-square analysis to determine the percentage of promoters from genes assigned to functionally related groups in *Aspergillus* species that were significantly enriched for TF binding sites involved in plant polysaccharide degradation.

Chapter 3

Phenotypic analyses of *sltA1* mutant strains

3.1 Introduction

3.1.1 The *sltA1* mutation of *A. nidulans*

O'Neil *et al.* (2002) described the cloning of a gene encoding a C_2H_2 zinc finger protein (StzA) that alleviated sensitivity to a variety of abiotic stresses in an *A. nidulans sltA1* mutant. The *sltA1* mutation was originally recognised as conferring sensitivity to Na^+ and K^+ (Spathas, 1978). Since then, complementation of the *sltA1* mutation with the *stzA* gene has indicated that this gene alleviates sensitivity to UV, the alkylating agent MNNG and high arginine concentrations when present as a nitrogen source, in addition to cations (Clement *et al*., 1996; O'Neil *et al.*, 2002).

Additionally, it was claimed that the *sltA1* mutant strain GO281 has a reduced ability to utilise glycerol and ethanol as sole carbon sources when compared to the wild-type strain L20 (O'Neil *et al*., 2002). Nonetheless, it was evident that complementation of the *sltA1* mutation by the cloned *stzA* gene in strains STC1−4 failed to restore wild-type growth on glycerol and ethanol (O'Neil *et al*., 2002; Barham-Morris, 2006). This indicated that *stzA* was not involved in the metabolism of these carbon sources. Furthermore, use of the cloned *stzA* gene to complement the *sltA1* mutation in strain WV202 (constructed by Clement *et al*., 1996) also failed to restore defects in glycerol and ethanol utilisation (Barham-Morris, 2006). Hence, in order to expand understanding of the pleiotropic nature of StzA regulation, the effect of the *sltA1* mutation upon a wide variety of carbon and nitrogen sources was analysed in the present study. Verification that these altered phenotypes were due to the *sltA1* mutation and not due to a separate uncharacterised mutation was important.

3.1.2 Carbon and nitrogen metabolism

Filamentous fungi possess the ability to utilise a diverse array of carbon and nitrogen sources by rearranging their metabolism to adapt to changes in substrate availability (reviewed in Fleck *et al*., 2011). Alterations in available carbon and nitrogen substrates lead to changes in the induction of specific enzymes for the breakdown of the particular compounds present and rearrangement of the flow of metabolites through central metabolic pathways. Such metabolic capabilities are not only relevant to saprophytic fungi adapting

to the soil environment but also to pathogenic fungi that must adapt to animal or plant host environments upon infection (Archer and Dyer, 2004). Global co-ordinating regulatory systems for the control of the utilisation of carbon and nitrogen sources have been studied in detail and are well documented (reviewed in Ruijter and Visser, 1997; Marzluf, 1997; Gancedo, 1998).

Carbon metabolism has been the subject of intense research within a broad spectrum of fungi, including *A. nidulans* (Ruijter and Visser, 1997; Gancedo, 1998). Carbon repression has been divided into three broad groups based on metabolic function (Ruijter and Visser, 1997). These were systems catabolising less favourable carbon sources (such as ethanol, amino acids and pectin), carbon repression of secondary metabolism, and gluconeogenic and glyoxylate cycle enzymes. The first of these groups, encompassed by the term carbon catabolite repression (CCR), is by far the most extensively studied of the three categories in *A. nidulans* and other filamentous fungi. The presence of a high concentration of a preferred carbon source, such as D-glucose, is required to enable the CCR response (Dowzer and Kelly, 1991). CCR is mediated through the transcription factor CreA, which represses genes required for the utilisation of less preferred carbon sources to favour glycolysis over the reverse, and more metabolically costly, process of gluconeogenesis. CreA binds to the consensus sequence 5'-SYGGRG in target promoters *via* two zinc fingers located in its N-terminal region (Cubero and Scazzocchio, 1994). Functional regulatory target sequences for CreA have been identified in the promoters of many genes, including *alcR*, the pathway specific activator for alcohol dehydrogenase I and aldehyde dehydrogenase (Felenbok *et al*., 2001); the xylanase gene, *xlnA* (Tamayo *et al*., 2008); and the proline permease gene, *prnB* (Gonzalez *et al*., 1997). CreB, CreC and CreD also participate in the regulation of carbon source utilisation, with CreD influencing the interaction between CreB and CreC (Lockington and Kelly, 2002; Kelly and Boase, 2004).

Given the relative abundance of information regarding CreA, it is perhaps surprising that less is known regarding the regulation of the other main systems of secondary metabolism and gluconeogenic and glyoxylate cycle enzymes in fungi. In *A. nidulans*, the genes *acuK* and *acuM* are known to encode transcriptional activators responsible for the regulation of gluconeogenesis, the process in which glucose is synthesised from non-carbohydrate carbon substrates (Hynes *et al*., 2007; Suzuki *et al*., 2012). Additionally, phosphoenolpyruvate carboxykinase (PEPCK), a key gluconeogenic enzyme, has been characterised in this organism (Hynes *et al*. 2002). The glyoxylate cycle is required for growth on gluconeogenic sources metabolised *via* acetyl CoA (Lorenz and Fink, 2001). Several studies have shown reduced virulence of

fungal strains lacking key glyoxylate cycle enzymes (isocitrate lyase and/or malate synthase), including for the animal pathogens *A. fumigatus* (Olivas *et al*., 2008) and *Candida albicans* (Lorenz and Fink, 2001) and also for the plant pathogens *Leptosphaeria maculans* (Idnurm and Howlett, 2002) and *Stagonospora nodorum* (Solomon *et al*., 2004).

The utilisation of nitrogen sources in *A. nidulans* is under the control of AreA, a positive-acting transcription factor that binds to the consensus sequence 5'-HGATAR (Wilson and Arst, 1998). In this manner, AreA regulates permeases and catabolic enzymes required for the utilisation of alternative, metabolically costly nitrogen sources (such as nitrate, amino acids and nucleotides) when more favourable nitrogen sourses (such as ammonia) are limiting. AreB negatively regulates nitrogen catabolism under nitrogen-limiting and nitrogen-starvation conditions (Wong *et al*., 2009). Compounds such as amino acids and acetamide can act as both a nitrogen and carbon source and have been found to be subject to well-integrated mechanisms of nitrogen and carbon control (Gonzalez *et al*., 1997; Macios *et al*., 2012). CpcA is the central regulator of the cross-pathway control system and responds to amino acid availability (Hoffmann *et al*., 2001). Ultimately, these mechanisms are co-ordinated to allow the best use of available compounds as carbon and nitrogen sources (Marzluf, 1997).

3.1.3 Sterigmatocystin and aflatoxins

Poor utilisation of the polyamine putrescine by GO281 was considered to imply possible disrupted sterigmatocystin production in this strain since polyamines are required for sterigmatocystin synthesis (Walters *et al*., 1997). Sterigmatocystin is a toxic and carcinogenic secondary metabolite produced by several species of the fungal genus *Aspergillus*, including *A. nidulans* (Dezotti and Zucchi, 2001). Moreover, sterigmatocystin is the penultimate precursor of aflatoxins in the aflatoxin biosynthetic pathway (Dezotti and Zucchi, 2001; Cary *et al*., 2005). Aflatoxin contamination of agricultural crops by *A. flavus* and *A. parasiticus* pose a significant risk to livestock and human health due to their toxicity and carcinogenicity (Yu *et al*., 2004a). As there is considerable homology in the organisation of the aflatoxin/sterigmatocystin biosynthetic genes among *A. flavus*, *A. parasiticus* and *A. nidulans* (Ehrlich *et al*., 2005), further understanding of the sterigmatocystin biosynthetic regulatory process in *A. nidulans*, which is non-pathogenic, may contribute to strategies for the control of aflatoxin contamination (Ehrlich *et al*., 2003; Yu *et al*., 2004a).

3.1.4 Aims

Growth tests were performed on solid media to analyse any phenotypic differences between the *sltA1* mutant strain GO281 and L20 (*sltA*$^{+}$) when grown on a range of carbon and nitrogen sources. Sterigmatocystin production between the two strains was also analysed. Sexual crosses were set up to determine whether the observed phenotypic differences could be attributed to the *sltA1* mutation.

3.2 Results

3.2.1 Simple and complex carbon source utilisation by L20 and GO281

Numerous carbon and nitrogen sources were analysed for their potential to be utilised by the *A. nidulans* strains GO281 (*sltA1*) and L20 (*sltA*$^{+}$). These results are summarized in Table 3.1 and representative examples of growth morphologies are shown in Figure 3.1. In conditions assessing growth on various carbon sources (1% w/v), ammonium tartrate (10 mM) was used as the sole nitrogen source because ammonium is a highly preferential nitrogen source for *A. nidulans*, allowing strong growth. When assessing growth on various nitrogen sources (10 mM), glucose (1% w/v) was used as the sole carbon source because it is a highly preferential carbon source (Marzluf, 1997; Ruijter and Visser, 1997).

A variety of simple compounds that act purely as carbon sources were initially analysed. The utilisation of glucose, sucrose, lactose, fructose, mannitol, sorbitol, arabinose, ribose and trehalose were all considered to be unaffected by the *sltA1* mutation (Table 3.1). Growth of GO281 on erythritol and arabitol was arrested to a small degree compared to L20. The utilisation of glycerol, galactose, galacturonic acid, glucuronic acid, xylose, rhamnose, acetic acid, quinic acid, benzoate, ethanol, erythrose, dihydroxyacetone and acetaldehyde was severely affected in GO281.

Table 3.1. Growth of *A. nidulans* L20 (*sltA*$^{+}$) and GO281 (s*ltA1*) on MM supplemented with sole carbon and nitrogen sources. Carbon sources were provided at a concentration of 1% and nitrogen sources at a concentration of 10 mM. When carbon source utilisation was investigated, ammonium tartrate (10 mM) was provided as a nitrogen source. Cellulose was the sole exception, where ammonium tartrate was replaced with 10 mM urea. When nitrogen source utilisation was investigated, glucose (1%) was provided as a carbon source. Key: +++ strong growth, ++ weaker growth, + no/poor growth.

Compound	L20	GO281	Compound	L20	GO281
Simple carbon sources					
Glucose	+++	+++	Galactose	+++	+
Sucrose	+++	+++	Galacturonic acid	+++	+
Lactose	+++	+++	Glucuronic acid	+++	+
Fructose	+++	+++	Xylose	+++	+
Mannitol	+++	+++	Rhamnose	+++	+
Sorbitol	+++	+++	Acetic acid	+++	+
Arabinose	+++	+++	Quinic acid	+++	+
Ribose	+++	+++	Benzoate	+++	+
Trehalose	+++	+++	Ethanol	++	+
Erythritol	+++	++	Erythrose	++	+
Arabitol	+++	++	Dihydroxyacetone	++	+
Glycerol	+++	+	Acetaldehyde	++	+
Complex carbon sources					
Xylan	+++	+++	Cellulose	++	+
Pectin	+++	+++	Oleic acid	++	+
Polygalacturonic acid (pH 6.5)	+++	+	Polygalacturonic acid (pH 8.5)	++	+
Starch	+++	+++			
Combined carbon and nitrogen sources					
Alanine	+++	+	Histidine	++	+
Arginine	+++	+	Leucine	++	+
Asparagine	+++	+	Lysine	++	+
Glutamic acid	+++	+	Methionine	++	+
Glycine	+++	+	Proline	++	+
Isoleucine	+++	+	Threonine	++	+
Phenylalanine	+++	+	Tryptophan	++	+
Serine	+++	+	Tyrosine	++	+
Aspartic acid	++	+	Valine	++	+
Glutamine	++	+	Cysteine	+	+
Putrescine	+++	+	Milk	+++	+++
Compounds of the TCA cycle					
Citric acid	++	+	Fumaric acid	++	+
Isocitric acid	++	+	Malic acid	++	+
Pyruvic acid	++	+	Oxoglutaric acid	++	+
Succinic acid	++	+	Oxaloacetic acid	+	+
Nitrogen sources					
Ammonium tartrate	+++	+++	Arginine (10 mM)	+++	+++
Sodium nitrate	+++	+++	Arginine (50 mM)	+++	+
Sodium nitrite	+++	+++	Proline	+++	+++
Glutamic acid	+++	+++	Phenylalanine	++	++
Glutamine	+++	+++	Urea	+++	+++
Glycine	+++	+++	Hypoxanthine	+++	+++

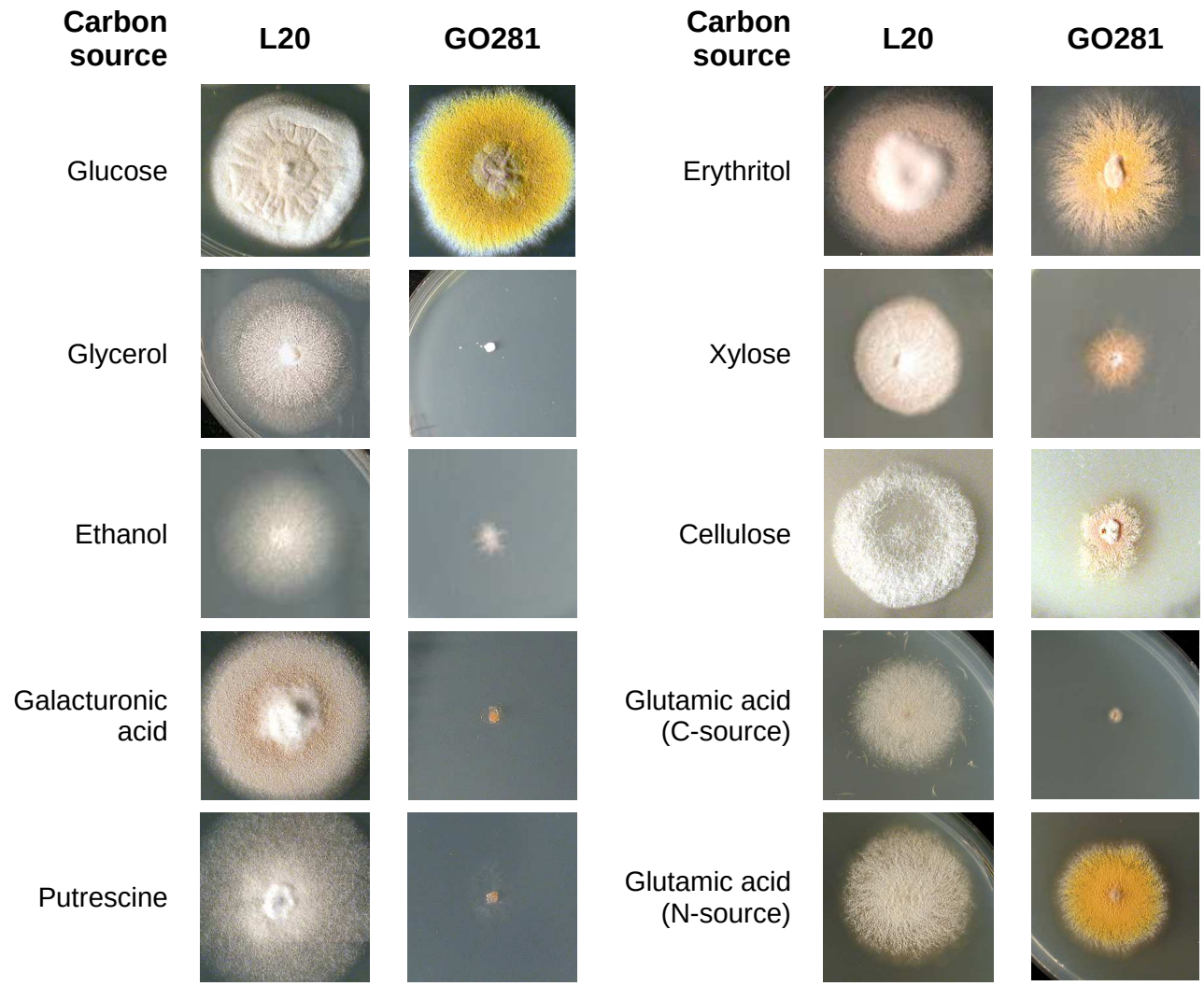


Figure 3.1 Examples of the growth morphologies of L20 (*sltA*$^{+}$) and GO281 (s*ltA1*) on different sole carbon sources. All colonies were incubated at 37 ºC for 72 hours.

Given that in *T. reesei*, the metabolism of xylan and cellulose was severely affected by the deletion of the *stzA* orthologue *ace1* (Aro *et al.,* 2003), the utilisation of complex polysaccharides by GO281 was compared with L20. Table 3.1 shows that utilisation of xylan and pectin was unaffected by the *sltA1* mutation, which did, however, confer a poor ability to use polygalacturonic acid. Growth of *A. nidulans* on cellulose was tested in the presence of urea as a nitrogen source since ammonia acts to inhibit cellulose utilisation (Lockington *et al*., 2002). Under these conditions, growth of GO281 was greatly reduced compared to L20 (Figure 3.1). Oleic acid utilisation was also reduced in GO281 compared to L20. Growth on starch was similar between the two strains and characteristic zones of clearing indicated the efficient production and secretion of amylase.

3.2.2 Utilisation of amino acids and TCA cycle intermediates

L20 grew well on all twenty amino acids, with the sole exception of cysteine, when used as sole carbon sources in the presence of 10 mM ammonium tartrate. In contrast, reduced growth was observed for the *sltA1* mutant GO281 on all the amino acids (Table 3.1). When grown on putrescine supplemented with ammonium tartrate, the growth of L20 was strong, while GO281 growth was poor (Figure 3.1). Strong growth of both L20 and GO281 was observed on milk as a sole carbon source.

The growth of L20 and GO281 on tricarboxylic acid (TCA) cycle intermediates used as sole carbon sources was analysed. L20 grew well on citric acid, isocitric acid, pyruvic acid, succinic acid, fumaric acid, malic acid, and oxoglutaric acid, although growth was very poor on oxaloacetic acid (Table 3.1). In contrast, GO281 grew poorly on all of these metabolites central to the TCA cycle.

3.2.3 Nitrogen source utilisation

Both L20 and GO281 grew well on all the nitrogen sources tested (10 mM): ammonium tartrate, sodium nitrate, sodium nitrite, glumatic acid, glutamine, urea, hypoxanathine, glycine, proline, phenylalanine and arginine (Table 3.1). Clement *et al*. (1996) reported that although *sltA1* mutants grew well on arginine as a sole nitrogen source, they demonstrated sensitivity to this amino acid at high concentrations (>10 mM). This was confirmed by the results of the present study (Table 3.1). No similar sensitivity of GO281 to high concentrations (50 mM) of other amino acids as sole nitrogen sources was observed in the present study (results not shown).

3.2.4 Sterigmatocystin assay using Thin Layer Chromotography

Thin Layer Chromotography (TLC) analysis was used to estimate the relative sterigmatocystin concentrations produced by L20 and GO281 and compared with known concentrations of sterigmatocystin (Figure 3.2). Estimated sterigmatocystin concentrations for L20 (*sltA*$^{+}$) and GO281 (*sltA1*) samples were 10 ng µl^{-1} and <0.2 ng µl^{-1}, respectively, indicating at least a 50-fold reduction in sterigmatocystin production by GO281 compared to L20 (Table 3.2).

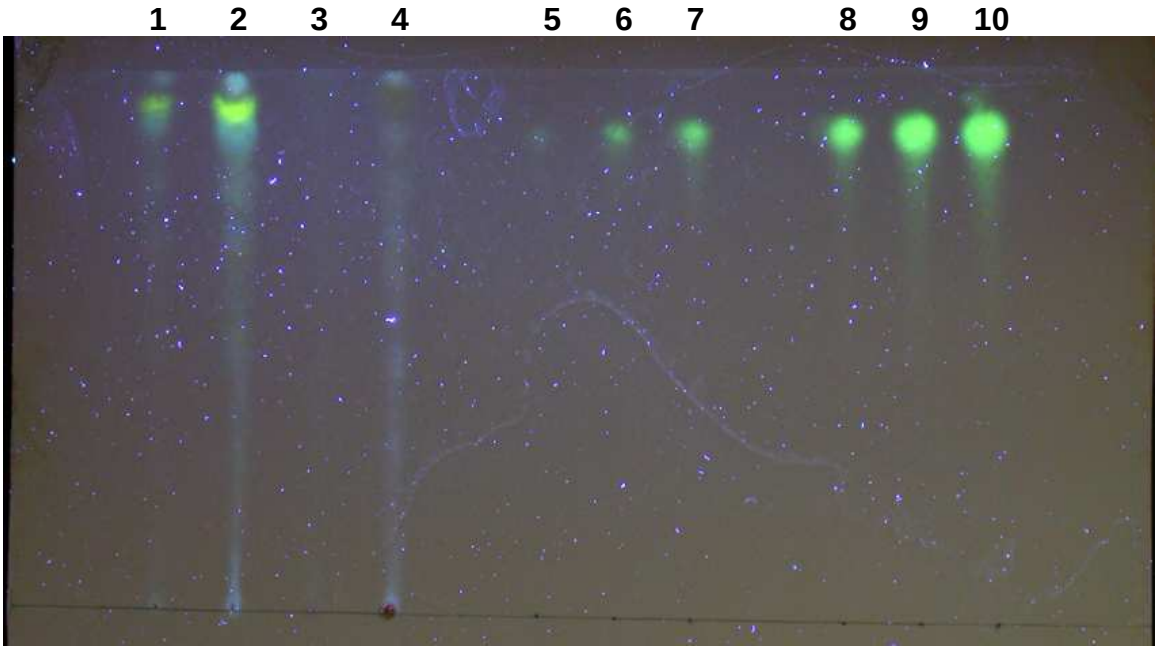


Figure 3.2. TLC plate showing stergimatocystin produced by L20 and GO281. Lanes 1 and 2 show sterigmatocystin produced by *A. nidulans* L20 (*sltA*$^{+}$) and lanes 3 and 4 show the amount produced by GO281 (*sltA1*). Lanes 5 to 10 show control concentrations of sterigmatocystin (5 ng, 25 ng, 50 ng, 100 ng, 500 ng and 1000 ng).

Table 3.2. Estimated concentrations of sterigmatocystin produced by L20 and GO281.

Lane	1	2	3	4	5	6	7	8	9	10
Sample	L20		GO281		Sterigmatocystin standards					
Sterigmatocystin concentration	~10 ng µl^{-1}		<0.2 ng µl^{-1}		5 µg ml^{-1}			100 µg ml^{-1}		
Volume plated (µl)	5	25	5	25	1	5	10	1	5	10
Sterigmatocystin concentration plated (ng)	~40	~300	<5	<5	5	25	50	100	500	1000

3.2.5 Sexual crosses with L19 and L20; WP1 and GO281

It was important to check that the phenotypes relating to altered carbon metabolism in the *sltA1* strain GO281 were due to the *sltA1* mutation and not due to a separate uncharacterised mutation. Sexual crosses allowed recombinant progeny to be generated, which could be tested for segregation of carbon utilisation phenotypes with salt sensitivity, defective DNA repair and arginine sensitivity phenotypes known to be affected in *sltA1* mutants. GO281 (*yA2 pabaA1*) was crossed with WP1 (*wA3 pyroA4*), a recombinant resulting from an earlier cross between L20 (*wA3 pabaA1*) and L19 (*yA2 pyroA4*). The resulting progeny were analysed phenotypically.

From the cross between GO281 and WP1, spore colours of the F1 progeny were scored (174 yellow and 167 white) and demonstrated conventional (1:1) segregation patterns (i.e. Mendelian inheritance). One hundred of the progeny (50 white and 50 yellow) were then randomly selected for further phenotypic analysis. Segregation of vitamin auxotrophies (para-aminobenzoic acid- and pyridoxine-requiring phenotypes) were initially analysed as a control. This analysis was performed using plate growth tests with MM containing glucose with the presence or absence of the vitamins para-aminobenzoic acid and/or pyridoxine. Colonies were incubated at 37 ºC for 48 hours. Analysis of the progeny revealed 1:1 segregation of the inheritance of the auxotrophic marker *pyroA4* and *pyroA*$^{+}$. It is unknown why there was a small but significant bias against *pabaA1* individuals. Table 3.3 summarises the inheritance of markers in the cross between GO281 and WP1.

Table 3.3. Inheritance of markers in the sexual cross between GO281 and WP1. The phenotypes relating to spore colour (*yA2* and *wA3*) and vitamin requirements (*pyroA4* and *pabaA1*) are known to be due to single mutations in *A. nidulans* genes (Clutterbuck, 1997). Hence, these alleles could be tested for 1:1 segregation patterns in the progeny resulting from a cross between GO281 and WP1 following sexual recombination. The χ^2 critical value is 3.84; P = 0.05.

Marker	Linkage group	No. of progeny		χ^2
Spore colour	I (*yA2*) II (*wA3*)	*yA2*	174	0.14
		wA3	167	
Pyro requirement	IV	*pyroA4*	49	0.04
		pyroA^{+}	51	
Paba requirement	I	*pabaA1*	38	5.76
		pabaA^{+}	62	
Altered carbon utilisation	Unknown	*gly*$^{-}$ *eth*$^{-}$ *gal*$^{-}$	42	2.56
		gly^{+} *eth*$^{+}$ *gal*$^{+}$	58	
Salt / DNA damage / arginine sensitivity	VI	*sltA1*$^{-}$	35	9.00
		sltA^{+}	65	

Plate tests were then used to test segregation of carbon utilisation phenotypes (glycerol, ethanol and galacturonic acid) with phenotypes of salt sensitivity (0.5 M NaCl), defective DNA repair (6 mins UV irradiation) and arginine sensitivity (50 mM arginine as a sole nitrogen source) attributed to *sltA1* mutants (Clement *et al*., 1996; O'Neil *et al*., 2002). The markers for altered carbon metabolism showed conventional (1:1) segregation patterns but those for *sltA1* did not. A stress resistant phenotype (*sltA*$^+$) was exhibited by 65% of the recombinant progeny compared to just 35% that were stress sensitive (*sltA1*). These phenotypes do not fit a 1:1 ratio because the calculated χ^2 value of 9.00 was much greater than the critical χ^2 value of 3.84 (P = 0.05). A similar bias against *sltA1* progeny was observed in the study by Clement *et al*. (1996), in which sexual progeny from a cross between GO281 and G051 (*sltA*$^+$) yielded only 32% *sltA1* individuals. Furthermore, in the same study, there were 10% fewer *pabaA1* individuals compared to *pabaA*$^+$ individuals, although not statistically significant. These results suggest a selective disadvantage of the *sltA1* and *pabaA1* genotypes at some stage during the sexual analysis, which may be related to a slower growth rate. The relative stronger growth of *sltA*$^+$ and *pabaA*$^+$ recombinant progeny on selective media compared to *sltA1* and *pabaA1* progeny may have led to the unintended but subjective preferential selection of larger colonies for further phenotypic analysis.

The precise phenotypes of the 100 recombinant progeny with regards to vitamin auxotrophies, *sltA1* and carbon utilisation phenotypes are summarised in Appendix 3. The growth morphologies of two of the progeny (D1 and D7) are shown in Figure 3.3, along with those of the parental strains GO281 and WP1 used as controls. The phenotypes of strains WV101 (*yA2 pyrG*$^+$ *sltA1*) and WV202 (*yA2 pyrG*$^+$ *sltA*$^+$) described by Clement *et al*. (1996, 1999) were also examined. The results demonstrated that mutant phenotypes associated with disrupted glycerol, ethanol and galacturonic acid utilisation always co-segregated, as did those associated with increased sensitivity to salt, UV and arginine. However, the altered carbon utilisation phenotypes segregated independently of salt, DNA repair and arginine sensitivity phenotypes. This indicates that the *sltA1* mutation in GO281 causes enhanced sensitivity to salt, UV and arginine and a separate mutation is responsible for the phenotypes relating to defective carbon source utilisation.

The *sltA1* mutation in *A. nidulans* WV101 was complemented with the *stzA* gene to create WV202 (Clement *et al*., 1996). Figure 3.3 shows that whilst WV202 (*sltA*$^+$) exhibited a wild-type response to NaCl, UV and arginine, it remained unable to utilise glycerol, ethanol and galacturonic acid in a similar manner to WV101

(*sltA1*). This is further convincing evidence that the altered carbon metabolism in these strains is due to a separate uncharacterised mutation that is unrelated to the *sltA1* mutation.

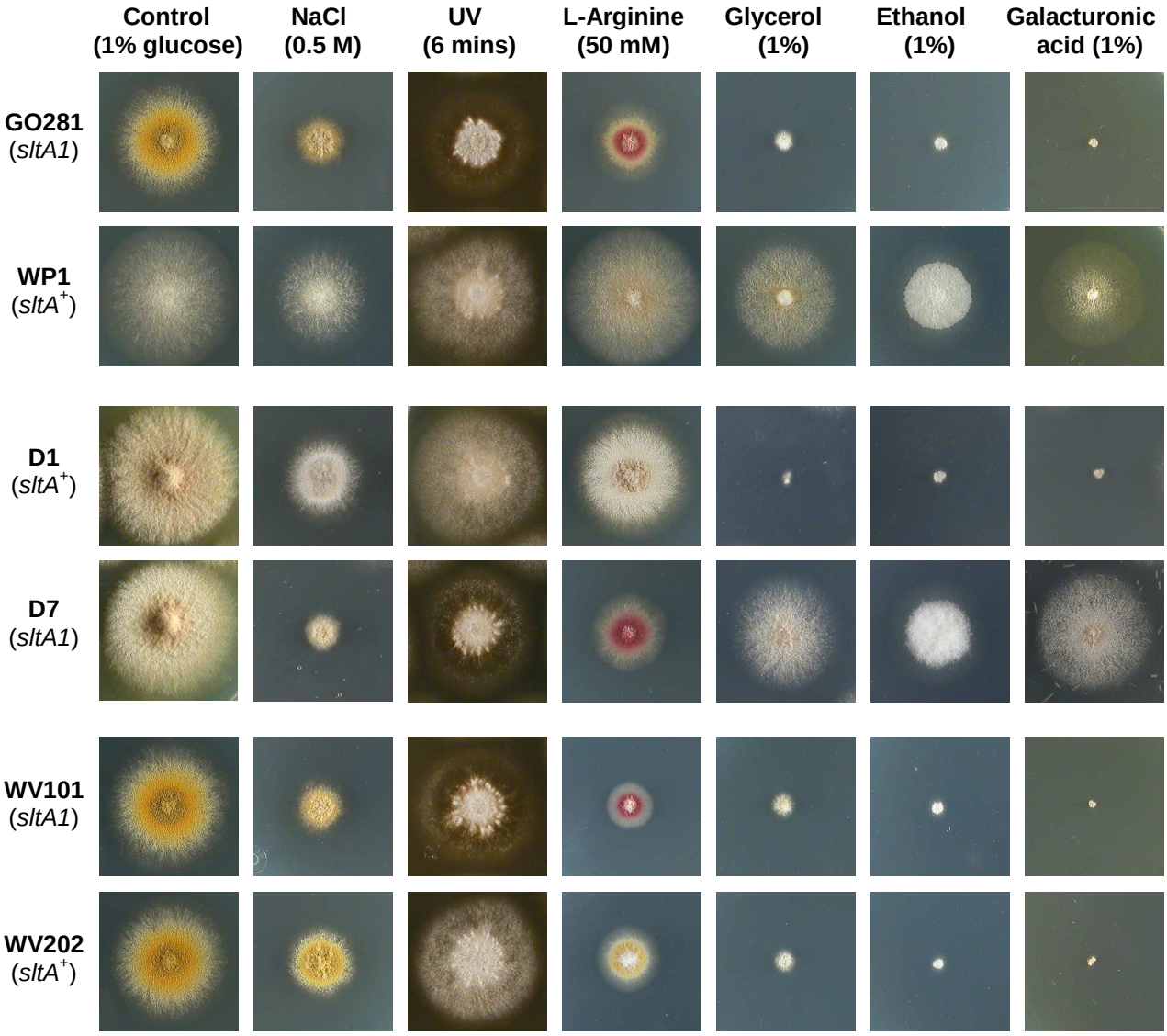


Figure 3.3. A collage of growth morphologies for *A. nidulans* strains subjected to differing stresses or differing sole carbon sources. Strains were incubated at 37 ºC for 48 hours on MM amended with either 0.5 M NaCl; L-arginine as a sole nitrogen source (50 mM); or glycerol, ethanol or galacturonic acid (1% w/v) used as sole carbon sources. UV stress was performed by irradiating colonies for 6 minutes after 24 hours of growth on MEA at 37 ºC, prior to incubation for a further 24 hours at 37 ºC. WP1 (*wA3 pyroA4*) represents a white *pyroA4* colony resulting from a cross between L20 (*wA3 pabaA1*) and L19 (*yA2 pyroA4*), and was subsequently crossed with GO281 (*yA2 pabaA1*). The growth morphologies of two selected recombinants (D1 and D7), resulting from the sexual cross between GO281 and WP1, demonstrated that defective carbon utilisation phenotypes segregated independently of *sltA1* phenotypes. WV101 (*yA2 pyrG*$^+$ *sltA1*) and WV202 (*yA2 pyrG*$^+$ *sltA*$^+$) are strains described by Clement *et al*. (1996, 1999). The results of the current study clearly demonstrate that although WV101 and WV202 behave differently with respect to responses to NaCl, UV irradiation and the presence of arginine (50 mM as a nitrogen source), both strains poorly utilise glycerol, ethanol and galacturonic acid when present as sole carbon sources. This indicates that defective carbon source utilisation is occurring independently of the *sltA1* mutation.

3.3 Discussion

Previously it has been shown that the *sltA1* mutation of *A. nidulans* GO281 affects a wide variety of phenotypes, including a reduced tolerance to a number of abiotic stresses, including NaCl, KCl, chemical mutagens and UV light, and arginine sensitivity (Spathas, 1978; Clement *et al*., 1996; O'Neil *et al*., 2002). It was also claimed that GO281 has a reduced ability to utilise glycerol and ethanol as sole carbon sources (O'Neil *et al*., 2002). Hence, the growth of this strain was assessed in the presence of a range of sole carbon and nitrogen sources. Although the wild-type strain L20 was able to utilise a diverse range of carbon sources, GO281 grew poorly on 41 of the 60 carbon sources examined (Table 3.1 and Figure 3.1).

Many of the carbon sources that were poorly utilised by GO281 are associated with gluconeogenesis, the metabolic pathway that results in the generation of glucose from non-carbohydrate carbon substrates (reviewed in Flipphi *et al*., 2009). Glycerol, ethanol, oleic acid, TCA cycle intermediates and amino acids were all poorly utilised by GO281 and all are substrates for gluconeogenesis, when used as sole carbon sources, after first being converted to pyruvate. These results indicate that the actions of one or both of the two key enzymes involved in gluconeogenesis (phosphoenolpyruvate carboxykinase, PEPCK; and fructose-1,6-bisphosphatase, FBP) are disrupted in GO281. PEPCK, encoded by *acuF*, is necessary for growth on compounds metabolised *via* the TCA cycle, and FBP, encoded by *acuG*, is necessary for growth on all carbon sources requiring gluconeogenesis (Hynes *et al*., 2002). Perturbed growth of GO281 on gluconeogenic carbon sources not metabolised through the TCA cycle (e.g. glycerol) strongly indicates that FBP activity is disrupted in GO281.

In fungi, there is a direct link between gluconeogenesis and the glyoxylate cycle. The glyoxylate cycle converts acetyl-CoA (produced during growth on carbon sources such as acetate, ethanol, and fatty acids) to oxaloacetate used for gluconeogenesis (Hynes *et al*., 2006). Hence, the regulation of the principle enzymes of the glyoxylate cycle may be defective in GO281, namely isocitrate lyase and malate synthase, encoded by *acuD* and *acuE*, respectively. Several studies have shown that the glyoxylate cycle is important for full fungal virulence. Glyoxylate cycle enzymes are essential for full plant pathogenesis in both *L*. *maculans* (Idnurm and Howlett, 2002) and *S*. *nodorum* (Solomon *et al*., 2004). These enzymes also account for the fully virulent phenotype of *C*. *albicans* (Lorenz and Fink, 2001) and *A. fumigatus* (Olivas *et al*., 2008) in mice. Moreover,

isocitrate lyase has been pinpointed as a potential drug target for antifungal drug development (Ebel *et al*., 2006).

Gluconeogenic carbon sources metabolised *via* acetyl-CoA also require FacB, which activates genes specifically required for growth on acetate (Todd *et al*., 1998). Although FacB induces glyoxylate cycle enzymes, it does not induce PEPCK nor other genes of the gluconeogenic pathway (Hynes *et al*., 2002). Furthermore, *facB* deletion mutants are able grow normally on glycerol. The inability of GO281 to utilise glycerol is indicative of a more general defective transcriptional regulation of gluconeogenesis in this strain. This may be indicative of defective transcriptional regulation of gluconeogenesis by *acuK* and *acuM* in GO281 (Hynes *et al*., 2007; Salazar *et al*., 2009).

Growth of GO281 on glycolytic carbon sources, such as glucose, sucrose, fructose, mannitol and sorbitol was similar to L20 suggesting that the glycolytic pathway functions normally in GO281 (Table 3.1). However, GO281 was unable to utilise galactose, on which L20 was able to grow well. Galactose can enter glycolysis *via* two metabolic pathways: the Leloir pathway and an alternative pathway that proceeds *via* L-sorbose (Fekete *et al*., 2004; Christensen *et al*., 2011). Hence, both pathways appear to be disrupted in GO281, while the main glycolytic pathway appears fully functional.

The utilisation of a number of plant substrates as sole carbon sources was compared between L20 and GO281 (Table 3.1). Interestingly, growth of GO281 on polygalacturonic acid (pH 6.5 and 8.5) was considerably reduced compared to the growth of L20. Growth was observed at pH 6.5 and 8.5 to ensure that both polygalacturonase and pectin lyase were being analysed (Zhao *et al*., 2007). It is perhaps surprising that both strains grew equally well on pectin, which also has a central polygalacturonic acid core. Pectin, however, contains a variety of side chains from this central core (reviewed in De Vries and Visser, 2001). These so called “hairy regions” contain galactose and arabinose, which GO281 has been shown to utilise. Likewise, the central polymer of xylan (β-1,4-linked D-xylose) contains side chains made up of galactose and arabinose (De Vries and Visser, 2001), which may explain why GO281 was able to utilise xylan despite xylose utilisation being severely affected in this strain. Compared to L20, growth of GO281 was also reduced on cellulose and on the gluconeogenic plant substrate oleic acid. Both strains grew reasonably well on starch.

Both L20 and GO281 grew equally well on all nitrogen sources tested at a concentration of 10 mM (ammonium tartrate, sodium nitrate, sodium nitrite, glutamic acid, glutamine, urea, hypoxanathine, glycine, proline, phenylalanine and arginine; Table 3.1). Hence, the regulation of AreA, which positively regulates genes required for the utilisation of nitrogen sources (Wilson and Arst, 1998), seems unaffected by *sltA1*. Since L20 and GO281 grew equally well on sodium nitrate and sodium nitrite, it seems unlikely that the *sltA1* mutation affects the nitrogen assimilation pathway. This pathway in *A. nidulans* involves NirA-mediated induction of genes (including *niiA* and *niaD*) required for uptake and reduction of nitrate and nitrite, which are subsequently reduced to ammonia before converted to glutamate and glutamine (Burger *et al*., 1991). It has previously been shown that GO281 is sensitive to high concentrations of arginine (>10 mM) when used as a sole nitrogen source, a toxic response characterised by decreased radial growth rate, reduced sporulation and production of a diffusible red pigment (Clement *et al*., 1996). This observation was supported by the results of the present study (Table 3.1 and Figure 3.3). GO281 was found not to be sensitive to other amino acids used as sole nitrogen sources at a concentration of 50 mM.

Poor utilisation of the polyamine putrescine by GO281 was considered to imply possible disrupted sterigmatocystin production in this strain since polyamines are required for sterigmatocystin synthesis (Table 3.1 and Figure 3.1; Walters *et al*., 1997). TLC analysis showed that GO281 is substantially impaired in sterigmatocystin production compared to L20 (Figure 3.2 and Table 3.2). However, a more recent analysis of sterigmatocystin production between *stzA*$^{+}$ and *stzA*$^{-}$ strains has revealed no observable differences in the concentrations produced, indicating that the *stzA* gene has no role in the regulation of sterigmatocystin synthesis (Kriskova, 2010). Nonetheless, as sterigmatocystin is the penultimate precursor of aflatoxins in the aflatoxin biosynthetic pathway (Dezotti and Zucchi, 2001; Cary *et al*., 2005), further studies of this process in *A. nidulans* strains such as GO281, which is non-pathogenic, may contribute to strategies for the control of aflatoxin contamination. Contamination of agricultural crops by aflatoxins from *A. flavus* and *A. parasiticus* currently pose significant risks to livestock and human health due to their toxicity and carcinogenicity (Ehrlich *et al*., 2003; Yu *et al*., 2004a).

Before it could be concluded that the *sltA1* mutation was responsible for the defective carbon source utilisation in GO281, it was necessary to verify this by analysing *sltA1* in a variety of genetic backgrounds. Phenotypic analysis of the recombinant progeny from a sexual cross between GO281 and WP1 (*sltA*$^{+}$) showed that simultaneous restoration of salt, UV and arginine sensitivity phenotypes was always observed

(i.e. *sltA1* phenotypes co-segregated). Likewise, simultaneous restoration of wild-type glycerol, ethanol, and galacturonic acid utilisation phenotypes always occurred. The *sltA1* sensitivity phenotypes, however, segregated independently of the altered carbon metabolism phenotypes (Figure 3.3 and Appendix 3). These results indicate that a separate uncharacterised mutation in GO281 is responsible for the phenotypes of defective carbon source utilisation, rather than the *sltA1* mutation. This uncharacterised mutation is not closely linked to the *sltA1* allele. A single uncharacterised mutation disrupting normal carbon metabolism in GO281 is supported by the 1:1 segregation pattern of wild-type and defective carbon utilisation phenotypes. Both WV101 (*sltA1*) and WV202 ($sltA^+$; Clement *et al*., 1996, 1999) poorly utilise glycerol, ethanol and galacturonic acid when present as sole carbon sources, although both strains behave differently with respect to responses to NaCl, UV irradiation and the presence of arginine. This is further evidence that perturbed carbon metabolism is occurring independently of the *sltA1* mutation in these strains.

3.4 Conclusion

Since GO281 has at least one uncharacterised mutation known to disrupt normal carbon metabolism, it now seemed necessary to create *stzA* deletion strains, rather than use GO281, to characterise the full range of phenotypes under StzA control.

Chapter 4

Construction and characterisation of *stzA* deletion strains

4.1 Introduction

4.1.1 Genetic manipulation of filamentous fungi

An increased understanding of fundamental genetic processes in model organisms has resulted in the development of new strategies for homologous recombination (HR) by DNA-mediated transformations (reviewed in Krappmann, 2007; Larrondo *et al*., 2009; Wortman *et al*., 2009; Kück and Hoff, 2010; Haas *et al*., 2011). These transformations are frequently used to create gene deletion strains. Until recently, DNA-mediated transformations in filamentous fungi have been difficult because self-replicating vectors are rare in filamentous fungi and the transferred DNA is often ectopically integrated into the genomic DNA (Kück and Hoff, 2010). Furthermore, DNA integration in filamentous fungi mainly occurs by non-homologous end-joining (NHEJ), and the site-specific recombination necessary for effective analysis of gene function occurs at very low frequencies (less than 1%). Hence, gene disruption experiments typically involved screening of hundreds of transgenic strains to identify the desired disruption strains.

Technical difficulties associated with DNA-mediated transformations in filamentous fungi have now been largely overcome by the introduction of Polymerase Chain Reaction (PCR)-based gene targeting. This involves using split-marker technology, which was initially developed for *S. cerevisiae* (Fairhead *et al*., 1996) but has since been successfully applied to *A. nidulans* (Nielsen *et al*., 2006; Nayak *et al*., 2006; Szewczyk *et al*., 2006) and other filamentous fungi (Krappmann, 2007; Larrondo *et al*., 2009). This split-marker technology involves fusion PCR, whereby marker gene fragments are fused with DNA fragments identical to those that flank the gene to be deleted. Two bipartite gene-targeting substrates are generated, which combine homologously with the target DNA region *in vivo* during transformation. During this process, the marker gene replaces the targeted gene.

The sequencing of the genomes of filamentous fungi has enabled gene sequences to be identified and their flanking regions to be amplified by PCR using primers based on genomic DNA sequences (Wortman *et al*., 2009). Hence, although the genes themselves are not amplified, their flanking sequences must be known. Since more than 100 fungal genome sequences have now been published, there is great potential for targeted gene manipulations to be performed to increase our understanding of gene functions in this group

of organisms containing species of medical and industrial importance (Kück and Hoff, 2010; Haas *et al*., 2011).

4.1.2 Advantages of bipartite-based gene targeting

The bipartite PCR-based gene-targeting method has several advantages compared to traditional gene deletion methods (Nielsen *et al*., 2006; Nayak *et al*., 2006). The gene-targeting substrate is synthesised solely by PCR, avoiding the need for several bacterial-based subcloning steps. Furthermore, the bipartite substrate is synthesised by two successive PCRs instead of the three required to construct continuous substrates in earlier methods (Yang *et al*., 2004; Yu *et al*., 2004b). Fewer PCRs save time and also reduce the possibility of undesired mutations occurring in the final gene-targeting substrates since even DNA proofreading DNA polymerases are not error free. Additionally, the presence of direct repeats at the ends of the *pyr4* sequences allows deletion of the *pyr4* selectable marker by the process of direct repeat recombination in the presence of 5-fluoro-orotic acid (5-FOA), enabling the marker to be recycled and subsequent genome manipulations to be performed more easily (Nielsen *et al*., 2006).

Gene targeting efficiencies for bipartite substrates were found to be threefold higher compared to continuous substrates, reflecting a higher ratio of correctly to ectopically integrated transformants that made the screening process to identify desired transformants easier (Nielsen *et al*., 2006). Furthermore, it was noted that the overall transformation efficiency with bipartite substrates was reduced compared to the efficiency obtained with continuous substrates. This was indicative of a reduced number of ectopic integrants with bipartite substrates since integration of the individual fragments of a bipartite substrate cannot produce viable transformants.

4.1.3 Advantages of using NHEJ recipient strains

The generation of gene knock-out mutants in filamentous fungi has previously been difficult due to the extremely low HR frequencies compared to *S. cerevisiae* (Nayak *et al*., 2006; Carvalho *et al*., 2010). Yu *et al*. (2004b) showed that successful gene targeting in *A. nidulans* was typically 20% or less for 24 out of 29 different fusion PCR products (with flanking regions ranging from 480 bp to 4.3 kb). Thus, although gene targeting was possible in *A. nidulans*, HR was very inefficient for wild-type strains.

Significant improvements in HR efficiencies could be achieved by using recipient strains that were deficient in DNA repair *via* non-homologous end-joining (NHEJ; Nayak *et al*., 2006; Krappmann, 2007). The NHEJ pathway is used for the repair of chromosomal DNA double-strand breaks (DSBs) within eukaryotes. This pathway does not require any homology and competes with the HR pathway (Shrivastav *et al*., 2008). The deletion of either component of the fungal Ku70–Ku80-protein complex (homologues of the human *ku70* and *ku80* genes), or alternatively the DNA ligase *lig4* gene, renders the NHEJ pathway inactive, favouring the HR of DNA sequences mediated by the Rad52 complex of the HR pathway (Figure 4.1). Thus, *ku* gene deletion strains allow targeted gene manipulations by preventing ectopic integration of gene fragments into the genome (Nayak *et al*., 2006; Krappmann, 2007). Moreover, the integration of one bipartite gene-targeting substrate by HR may help to channel the second substrate into the HR pathway (Nielsen *et al*., 2006).

Substantial increases in gene replacement frequencies have been reported for 14 species of filamentous fungi for which *ku* disruption strains have been constructed (Kück and Hoff, 2010). The disruption of either *ku70* or *ku80* generally resulted in significantly increased HR frequencies (typically between 50% and 100%) compared to less than 2% for wild-type strains. Although *ku* deletion dramatically improved gene targeting in *A. nidulans*, this deletion has no apparent effect upon growth or sensitivity to mutagens in this species (Nayak *et al*., 2006).

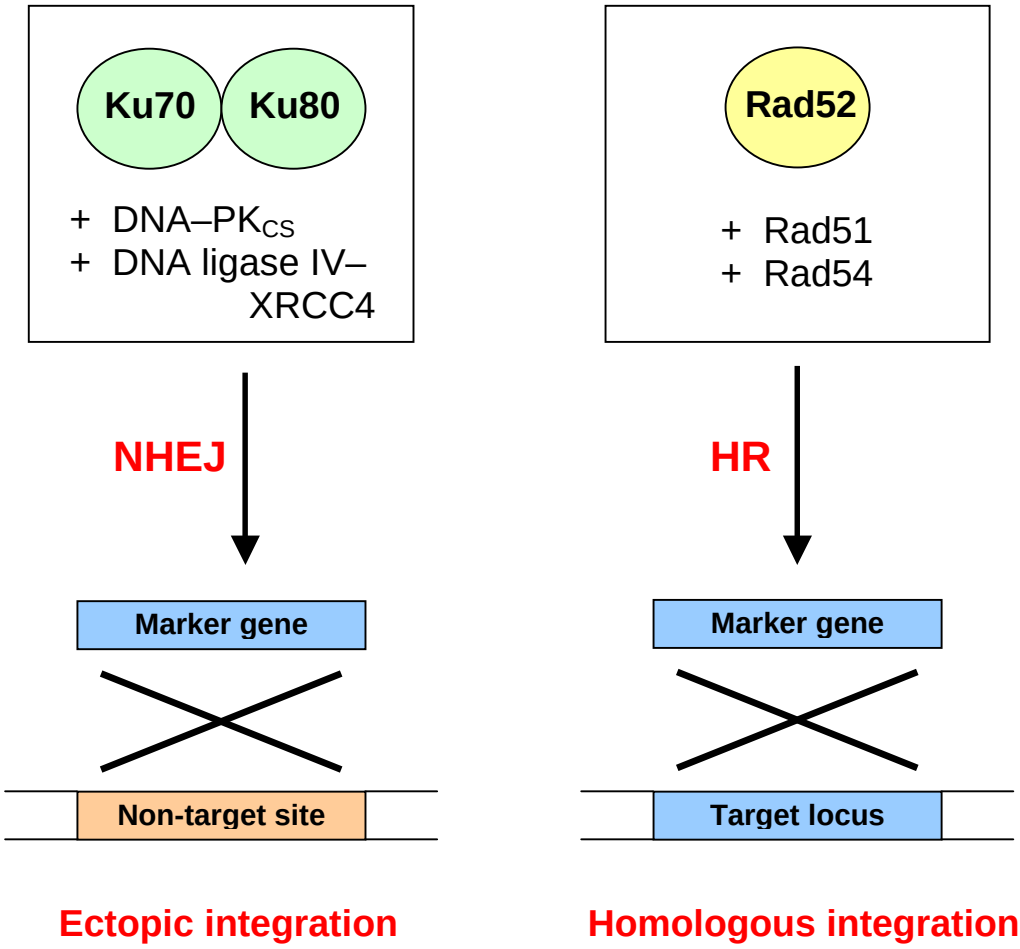


Figure 4.1. The NHEJ and HR pathways repair double-strand DNA breaks in fungi. Adapted from Krappmann (2007). The NHEJ (non-homologous end-joining) pathway is involved in the ectopic integration of DNA fragments and requires no homology. This pathway involves the Ku complex (Ku70–Ku80), which binds to the ends of double-stranded DNA in a non-specific fashion. The Ku complex recruits protein complexes such as the catalytic subunit of DNA-dependent kinase DNA–PK_{CS} and the XRCC4–DNA ligase IV complex. Competing with the NHEJ pathway is the HR (homologous recombination) pathway, which does require homology. In the HR pathway, the DNA ends interact with Rad52 and additional recombination proteins, such as Rad51 and Rad54, to promote HR.

4.1.4 A need for the construction of *stzA* deletion strains

The phenotypes controlled by StzA are not known with certainty. Confusion arises from the results of several studies. O'Neil *et al*. (2002) states that the *sltA1* mutation of *A. nidulans* results not only in enhanced sensitivity to salt, DNA-damaging agents and arginine but also in a reduced ability to utilise glycerol and ethanol as sole carbon sources. Nonetheless, it was evident that complementation of the *sltA1* mutation by the cloned *stzA* gene in strains STC1−4 failed to restore wild-type growth on glycerol and ethanol (O'Neil *et al*., 2002; Barham-Morris, 2006). This indicated that *stzA* was not involved in the metabolism of these carbon

sources. Furthermore, use of the cloned *stzA* gene to complement the *sltA1* mutation in strain WV202 (constructed by Clement *et al*., 1996) also failed to restore defects in glycerol and ethanol utilisation (Barham-Morris, 2006). Moreover, the *sltA1* strain GO281 was shown to be not only defective in the utilisation of glycerol and ethanol but also defective in the utilisation of a wide range of carbon sources (Chapter 3). The present study confirms that GO281 contains a secondary mutation affecting carbon metabolism that is not associated with *stzA* function. This evidence was derived from phenotypic analyses of recombinant progeny resulting from sexual crosses between *sltA1* and *sltA*$^{+}$ strains (Chapter 3). Hence, it seems likely that the *sltA1* strains GO281, WV101, WV202 and STC1−4 all possess at least one uncharacterised secondary mutation affecting carbon metabolism.

Other discrepancies arise from the study by O'Neil *et al*. (2002) in which phenotypes associated with the *sltA1* mutation in GO281 were not fully complemented by the *stzA* gene in strains STC1−4. Only STC1 and 2 were fully complemented with respect to salt, UV, 4NQO and MNNG sensitivities. STC3 did not exhibit complementation of MNNG sensitivity and STC4 did not exhibit complementation of salt, 4NQO or MNNG sensitivities. It is unknown whether different levels of *sltA1* complementation seen in STC1−4 were due to ectopic and multiple integration events that can reduce the apparent level of complementation of one or more mutations (Whitehead, 1990).

The construction of *stzA* deletion strains and subsequent phenotypic analyses were deemed to be necessary to ascertain precisely the phenotypes regulated by the *stzA* gene. At the onset of this work *stzA* = *sltA* had not been proven. The apparent complementation of the *sltA1* mutation in *A. nidulans* by *stzA* (O'Neil *et al*., 2002) may have been due to a gene dosage effect that altered gene transcription at a variety of loci, particularly as StzA is a transcription factor. Deletion of *stzA* to demonstrate a *sltA1* phenotype was a necessary step to imply allelism.

The *stzA* deletion strains originating from A1149 (A2, A5 and A7) and G191 (G6 and G25) were analysed for their growth responses on media imposing various conditions of stress and nutrition. This enabled a more precise determination of candidate genes controlled by StzA. Strains transformed only with the plasmid pDJB3 from A1149 (AP1 and AP2) and G191 (GP1 and GP2) were used as controls. Apart from *stzA* deletion and the differing loci of *pyr4* integration(s), strains A2, A5 and A7 were isogenic to AP1 and AP2. Similarly, G6 and G25 were isogenic to GP1 and GP2. Additionally, the growth phenotypes of L20, which is

wild-type with respect to *sltA*$^{+}$ and GO281, carrying the *sltA1* mutation, (and at least one secondary mutation affecting carbon metabolism; Chapter 3) were analysed.

4.1.5 *A. nidulans stzA* gene deletion methodology

It was decided to create deletion mutants for the *stzA* gene in *A. nidulans* using the split-marker system so that the full range of phenotypes controlled by StzA could be ascertained. This approach was deemed to be necessary since the *sltA1* mutant strain GO281 contains a secondary mutation affecting carbon metabolism, (Chapter 3) and *A. nidulans stzA* deletion mutants had yet to be created.

The gene-targeting system used to delete the *A. nidulans stzA* gene (AN2919) was based on the method of Nielsen *et al*. (2006) and was successfully used by that research group to delete the *radC* gene in *A. nidulans.* This system involves two bipartite gene-targeting substrates that both contain a targeting sequence and part of a selectable marker gene. However, neither substrate alone contains a functional marker. The marker is only generated following HR events involving the targeting sequences and overlapping marker sequences following co-transformation, thus allowing target specificity.

Two specific gene-targeting fragments flanking the *stzA* ORF were generated in two independent PCR reactions using *A. nidulans* L20 genomic DNA as a template (using primer pairs U1 and U2; D1 and D2; Figure 4.2). These primers pairs were designed by acquiring knowledge of the *A. nidulans* genomic DNA sequences flanking the *stzA* gene that resulted from the sequencing of the *A. nidulans* genome (Galagan *et al*., 2005) and could be accessed from the Broad Institute website (www.broadinstitute.org/).

The plasmid pDEL1 served as a PCR template for generation of the marker-containing fragments of the bipartite substrates because this vector contains the *N. crassa pyr4* gene, which can complement the defective *pyrG* gene in *A. nidulans* strains harbouring the *pyrG89* mutation (Oakley *et al*., 1987; Nielsen *et al*., 2006). The marker gene was generated in two independent reactions (using primer pairs R1 and M3; M2 and R2) that generated the ⅔ upstream and ⅔ downstream *pyr4* gene portions, respectively.

The fragments were fused together in pairs to create two bipartite gene-targeting substrates. The upstream *stzA*-targeting fragment (U) was fused with the fragment containing the upstream ⅔ portion of the *pyr4* gene (P1) using the nested primers F1 and F2. Similarly, in a separate reaction, the downstream *stzA*-targeting fragment (D) was fused with the fragment containing the downstream ⅔ portion of the *pyr4* gene (P2) using the nested primers F3 and F4. The specific fusion of two PCR-amplified fragments was possible since the fragments contained identical sequences in their ends, which acted as fusion tags. These tags were present at the end of the primers used to generate the initial fragments. Hence, the 3'-end acted as a normal PCR primer and the 5'-end contained a specific fusion tag; such oligonucleotides are often referred to as adaptamers (Erdeniz *et al*., 1997). During co-transformation, a functional marker was generated *in vivo* following HR of the fusion products consisting of targeting sequences and overlapping *pyr4* sequences. Thus, *pyr4* replaced the *stzA* ORF. The entire gene deletion process is summarised in Figure 4.2.

A. nidulans strains used for transformation were A1149 and G191. As both strains contain the *pyrG89* mutation, they were suitable recipients owing to their lack of functional *pyrG* genes (Oakley *et al*., 1987). As *pyrG* is essential for uracil biosynthesis, these strains normally require uridine and uracil for growth. However, *stzA* gene deletion transformants were able to grow on media lacking these nutrients due to the integration of the *N. crassa pyrG* orthologue *pyr4* at the *stzA* locus. The *A. nidulans ku80* deletion strain A1149 was used as one of the recipient strains for *stzA* deletion as strains defective in NHEJ pathways are excellent recipient strains for gene-targeting approaches (Kück and Hoff, 2010).

4.1.6 Aims

The aim was to construct and characterise *stzA* deletants to predict possible StzA target loci.

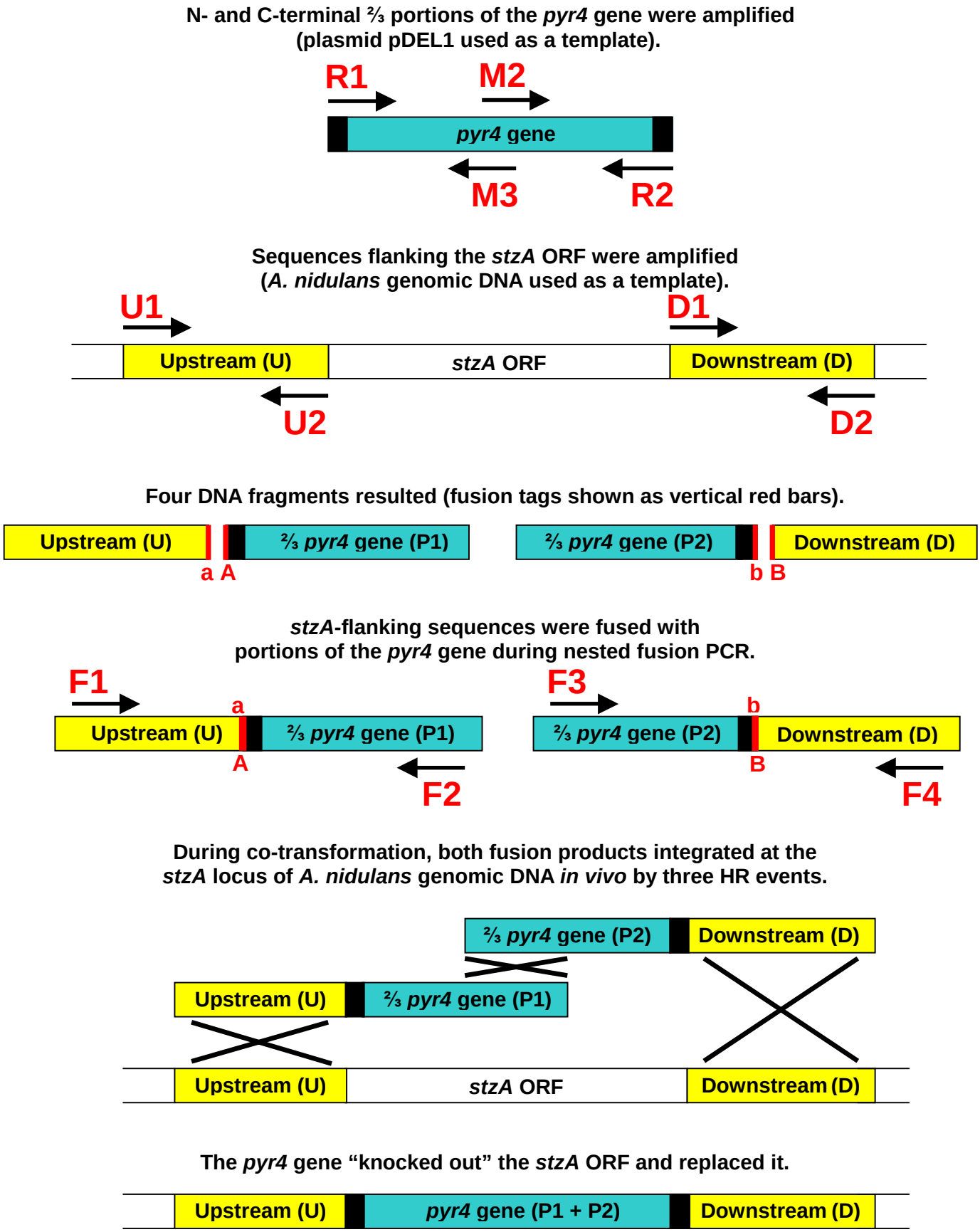


Figure 4.2. Strategy for *stzA* gene deletion. The positions of primers are indicated by arrows. Black boxes flanking the *pyr4* gene indicate direct repeats (292 bp). Fusion tags are represented by vertical red bars. The letters **a** and **A**; **b** and **B** represent complementary fusion tag sequences, but **A** and **B** are not related. Parallel bars represent the chromosome. Crosses indicate regions where HR occurs. For fusion PCR, only the primers used for the nested approach are shown (F1–F4). Fusion PCR was also attempted using different combinations of primer pairs used in the synthesis of the original DNA fragments (U1 and M3; M2 and D2).

4.2 Results: Construction of *stzA* deletion strains

4.2.1 Cloning of plasmids pDEL1 and pDJB3

Plasmid pDEL1 was constructed and kindly provided by Nielsen *et al*. (2006). It contains the *N. crassa pyr4* auxotrophic marker gene necessary to complement the *pyrG89* mutation in *A. nidulans*. *pyr4* complements this mutation by encoding orotidine-5'-phosphate decarboxylase, which is involved in the pyrimidine biosynthetic pathway, and is thus able to restore wild-type growth in the absence of uridine and uracil. Hence, pDEL1 was used as a template to amplify *pyr4* fragments to be used in fusion PCR reactions and to transform *A. nidulans pyrG89* strains. pDEL1 was cloned and then quantified by agarose gel electrophoresis by comparison to bands with known quantities of DNA obtained by restricting 1 µg of Lambda DNA with *Hin*dIII (Figure 4.3). The plasmid concentration was ~10 ng μl^{-1}. Successful restriction of the plasmid with *Hin*dIII indicated that the plasmid would be sufficiently pure for subsequent PCR.

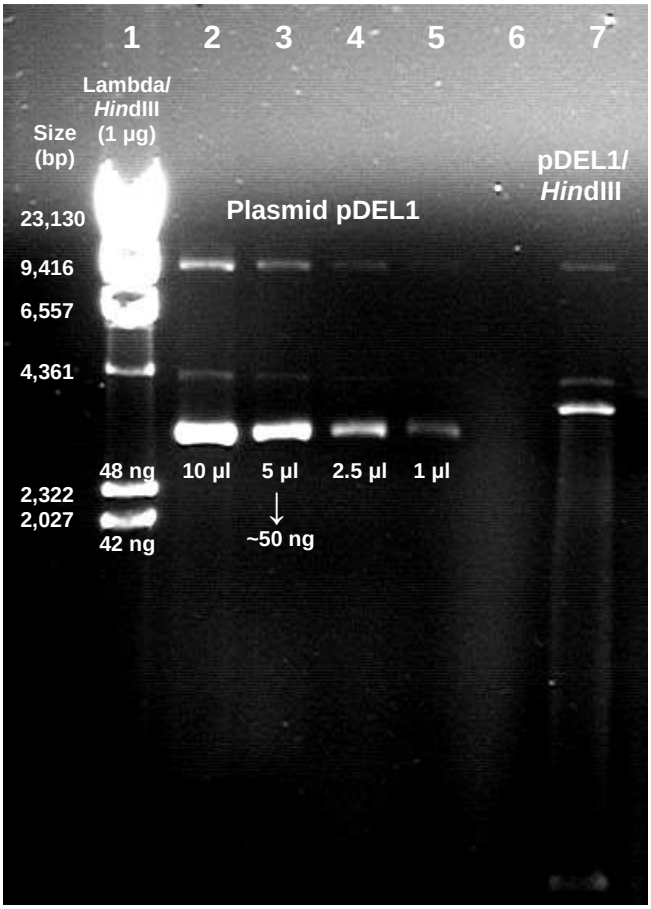


Figure 4.3. Purified plasmid pDEL1. Lane 1 shows a Lambda/*Hin*dIII digest (1 µg of restricted Lambda DNA). Lanes 2–5 show purified plasmid pDEL1 (10 µl, 5 µl, 2.5 µl and 1 µl of the sample). Lane 7 shows 5 µl of pDEL1 digested with *Hin*dIII.

The plasmid pDJB3, which also contained the *N. crassa pyr4* marker gene (Ballance and Turner, 1985) was also amplified and quantified (Figure 4.4). The plasmid concentration was ~0.5 µg µl^{-1}. Successful restriction of pDJB3 with *Hin*dIII and *Sal*I indicated that it would be sufficiently pure for subsequent PCR. pDJB3 was used to optimise transformation efficiency in the *A. nidulans pyrG89* strains G191 and A1149 before transforming them with the bipartite *stzA* gene-targeting substrates.

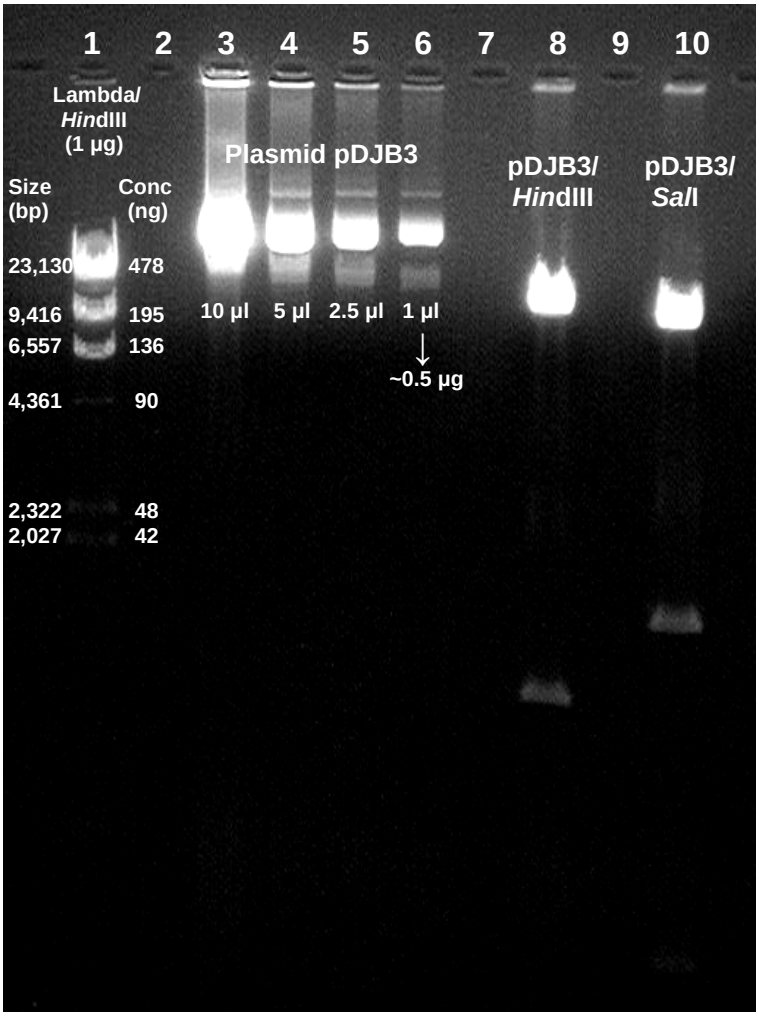


Figure 4.4. Purified plasmid pDJB3. Lane 1 shows a Lambda/*Hin*dIII digest (1 µg of restricted Lambda DNA). Lanes 3–6 show purified plasmid pDJB3 (10 µl, 5 µl, 2.5 µl and 1 µl of the sample). Lanes 8 and 10 show 5 µl of pDJB3 digested with *Hin*dIII and *Sal*I, respectively. A map of pDJB3 is shown in Appendix 4.

4.2.2 Isolation of genomic DNA from *A. nidulans* L20

Genomic DNA was isolated from *A. nidulans* L20 and quantified by comparison with bands of known DNA quantities in 5 µl of DNA Hyperladder I from Bioline (Figure 4.5). The genomic DNA concentration was ~16 ng μl^{-1}. A successful overnight restriction digest of the DNA with *Hin*dIII (Lane 6) indicated that the DNA sample was sufficiently pure to be used for PCR.

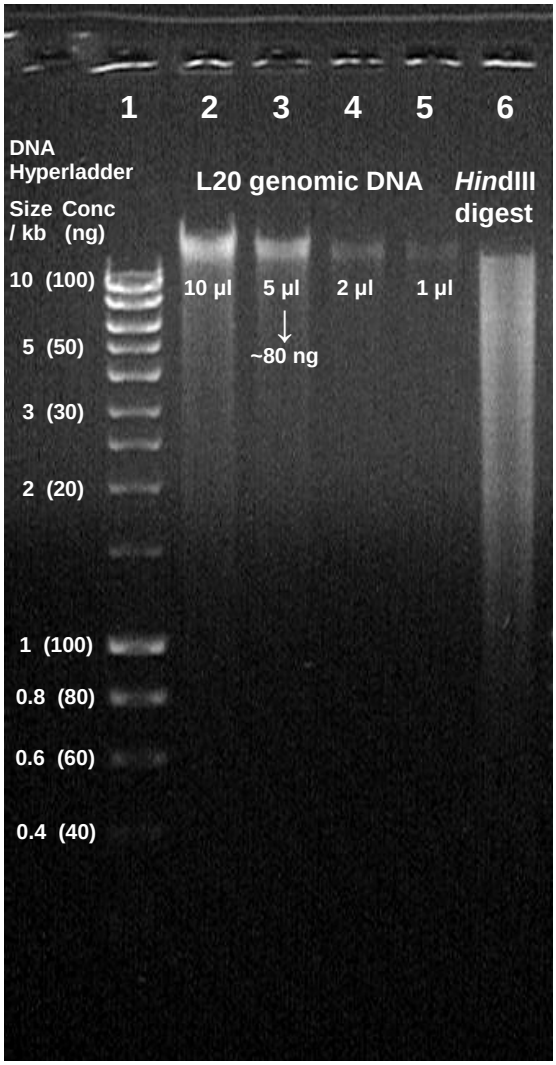


Figure 4.5. Purified genomic DNA from *A. nidulans* L20. Lane 1 shows 5 µl of DNA Hyperladder I from Bioline (720 ng). Lanes 2–5 show 10 µl, 5 µl, 2 µl and 1 µl of the genomic DNA sample. Lane 6 shows purified genomic DNA (30 µl) restricted with *Hin*dIII.

4.2.3 Primers used for amplifying *stzA*-flanking and *pyr4* gene fragments

The online Primer 3 program (http://frodo.wi.mit.edu/primer3/) was used to design primers to amplify both upstream (U) and downstream (D) regions immediately flanking the *stzA* ORF. These regions (990 and 822 bp) could be generated theoretically by PCR using the primer pairs U1 and U2; D1 and D2 (Table 4.1; Figure 4.6). Importantly, these regions were between 500 bp and 2 kb, a size range that has been shown to allow effective HR of gene-targeting substrates (Nayak *et al*., 2006). Note that primers U2 and D1 have additional sequences attached that act as fusion tags (shown in lower case).

The primers used for the amplification of the two *N. cras*sa *pyr4* gene portions were designed by Nielsen *et al*. (2006) and were present on the plasmid pDEL1 (Figure 4.7), which was constructed by the same research group. The N- and C-terminal ⅔ portions of the *pyr4* gene (designated P1 and P2, respectively) were amplified using the primer pairs R1 and M3; M2 and R2, respectively (Table 4.1). Primers R1 and R2 have additional sequences attached that act as fusion tags. R1 contains a fusion tag that is the reverse complement of that contained on U2. Similarly, R2 contains a fusion tag that is the reverse complement of that contained on D1.

Amplification of fusion products 1 and 2 was attempted using primer pairs U1 and M3; M2 and D2. Primers were also designed for the amplification of fusion products by nested PCR. This strategy involved amplification of fusion products 1 and 2 using primer pairs F1 and F2; and F3 and F4. All oligonucleotides used in this study were synthesised by MWG Biotech AG and are listed in Table 4.1.

Table 4.1. Oligonucleotides used to generate and fuse PCR fragments. Sequences in upper case represent sequences that function as primers, and sequences in lower case function as fusion tags. The fusion tags present on R1 and U2; R2 and D1 are complementary.

Name	Oligonucleotide sequence 5'–3'	Function
U1	CTCGCAGTTCAGCATCTTCA	Synthesis of *stzA*-flanking regions
U2	gatccccgggaattgccatgCCCTAATGTCCGAAAAGGTG	
D1	aattccagctgaccaccatgCTCGTCACATTTGGTGGTTG	
D2	TAGGCCATCAACGGATAAGG	
R1	catggcaattcccggggatcTGGATAACCGTATTACCGCC	Synthesis of *pyr4* selective marker fragments
R2	catggtggtcagctggaattTGCCAAGCTTAACGCGTACC	
M2	CCAGAAGCAGTACACGGC	
M3	GTTGTCTGCTTGCGCTTCTTC	
F1	ATGGTGGAGTTGCAGTTGG	Nested primers for fusion PCRs
F2	CTTGGTAATGCCGTTCAGG	
F3	ATCACCAACGCCGACATC	
F4	CGCTTGGAAGGCGAATATC	
S1	CCAGAAGCAGTACACGGC (same as M2)	PCR assay and sequencing
S2	GACAACGGCACTTTACAACG	
SA	GTTCTCTACGAAGCTCAACC	
SB	GTTTCTAGACAGATGTGTTCG	

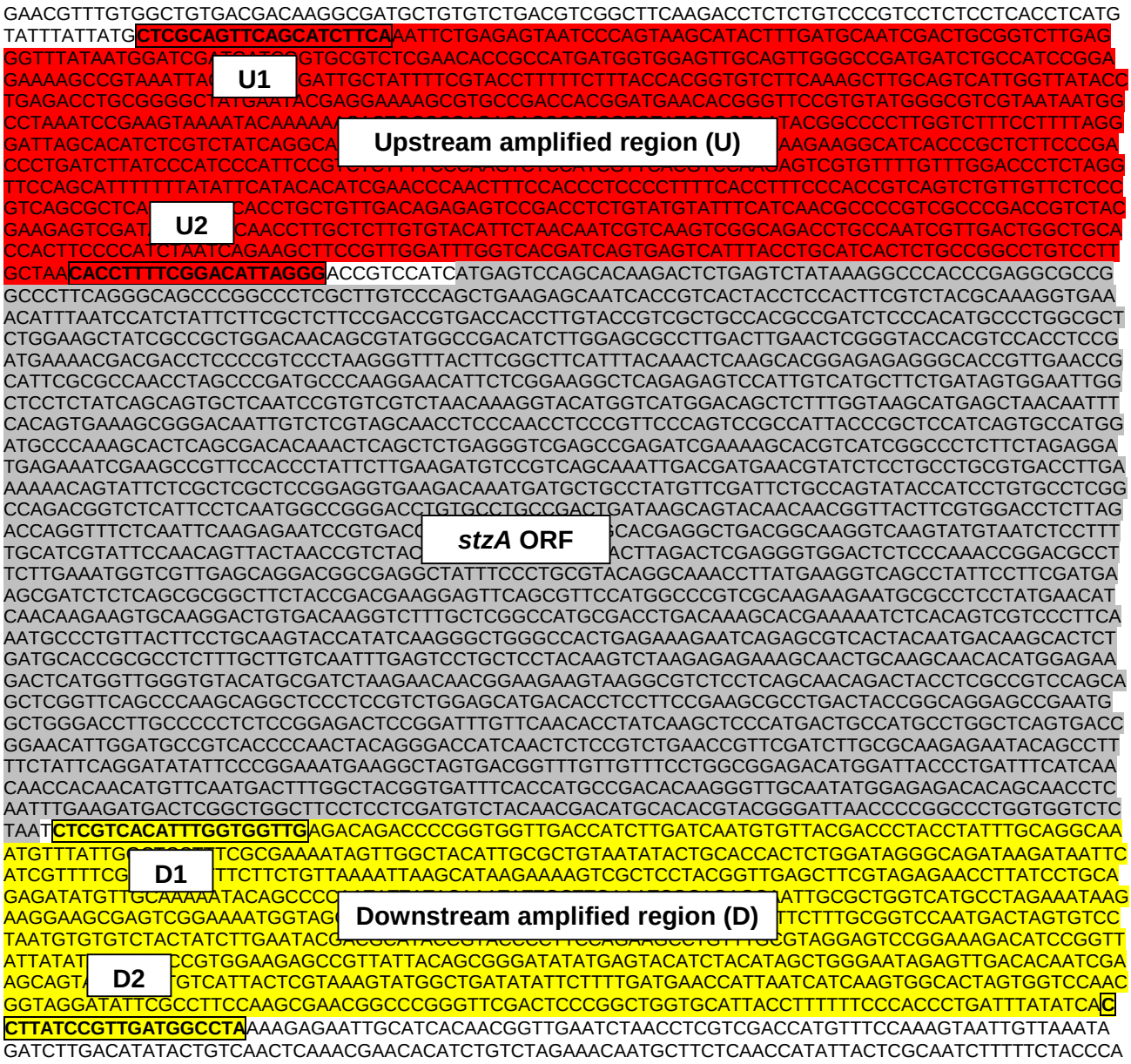


Figure 4.6. Nucleotide sequences flanking the *stzA* ORF. The ORF of the *stzA* gene (2,202 bp) is highlighted in grey (O'Neil *et al*., 2002). Upstream (U) and downstream (D) regions flanking the *stzA* ORF are highlighted in red and yellow. These regions are 990 and 822 bp and were synthesised using primer pairs U1 and U2; D1 and D2. Primer sequences are boxed and shown in bold.

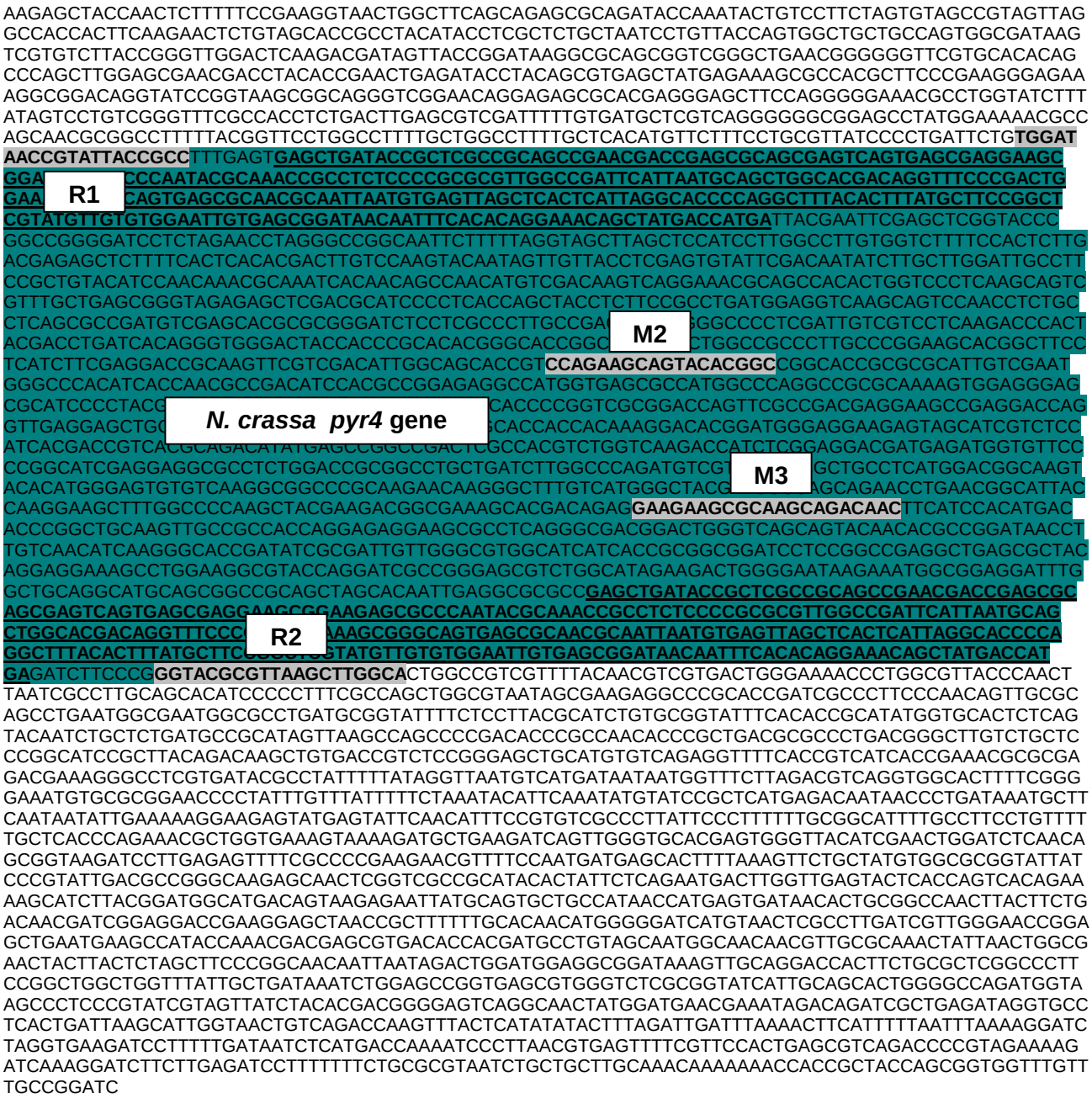


Figure 4.7. Plasmid pDEL1 nucleotide sequence. Plasmid pDEL1 was constructed by Nielsen *et al*. (2006). Nucleotides that comprise the *N. crassa pyr4* gene (along with flanking direct repeat sequences; 292 bp) are highlighted in green. The two overlapping ⅔ portions of the *pyr4* gene (P1 and P2) were generated using primer pairs R1 and M3; M2 and R2 (highlighted in grey). The 292 bp direct repeat sequences are shown in bold and underlined.

4.2.4 Amplification of *stzA*-flanking and *pyr4* gene fragments

The *stzA*-flanking and *pyr4* gene fragments were amplified by PCR using *A. nidulans* L20 genomic DNA and pDEL1 as templates. Very similar PCR reaction conditions and identical thermocycler conditions were used to synthesise each of the initial four DNA fragments (U, D, P1 and P2; Table 4.2). Phusion® polymerase (Finnzymes) was used for DNA amplification because it yielded blunt-ended DNA fragments that would not interfere with their subsequent fusion. This enzyme additionally possesses 3'–5' exonuclease proofreading activity to increase fidelity. Primer annealing temperatures were chosen as a compromise for each pair, with acceptable yields of all four fragments being obtained at 60 ºC. The amplified fragments were run on a 0.8% agarose gel and subsequently excised, purified and concentrated using a Qiagen DNA purification kit (Figure 4.8). The purified DNA fragments were each present at a concentration of ~100 ng μl^{-1}.

Table 4.2. PCR conditions used to amplify *stzA*-flanking and *pyr4* gene fragments.

(a) Reaction conditions

Component	Stock conc	Final conc	PCR-amplified fragment			
			U (990 bp)	**D** (822 bp)	**P1** (1536 bp)	**P2** (1270 bp)
SDW			127.4 µl	127.4 µl	142.4 µl	142.4 µl
5x HF buffer	5x	1x	40 µl	40 µl	40 µl	40 µl
$MgCl_2$ (In HF buffer)	7.5 mM	1.5 mM	-	-	-	-
dNTPs	25 mM	0.2 mM	1.6 µl	1.6 µl	1.6 µl	1.6 µl
Primer 1	50 µM	0.5 µM	U1 – 2 µl	D1 – 2 µl	R1 – 2 µl	R2 – 2 µl
Primer 2	50 µM	0.5 µM	U2 – 2 µl	D2 – 2 µl	M3 – 2 µl	M2 – 2 µl
Template DNA	L20 genomic: 16 ng μl^{-1} pDEL1: 10 ng μl^{-1}	L20 genomic: 2 ng μl^{-1} pDEL1: 0.5 ng μl^{-1}	L20 – 25 µl = 400 ng	L20 – 25 µl = 400 ng	pDEL1 – 10 µl = 100 ng	pDEL1 – 10 µl = 100 ng
Phusion® polymerase	2 U μl^{-1}	0.01 U μl^{-1}	2 µl	2 µl	2 µl	2 µl
Total volume			= 200 µl	= 200 µl	= 200 µl	= 200 µl

(b) Thermocycler conditions

Initial denaturing
98 ºC – 30 sec

Loop 1
40 cycles
98 ºC – 10 sec
60 ºC – 30 sec
72 ºC – 2.5 min

Final extension
72 ºC – 10 min

Final hold (4 ºC)

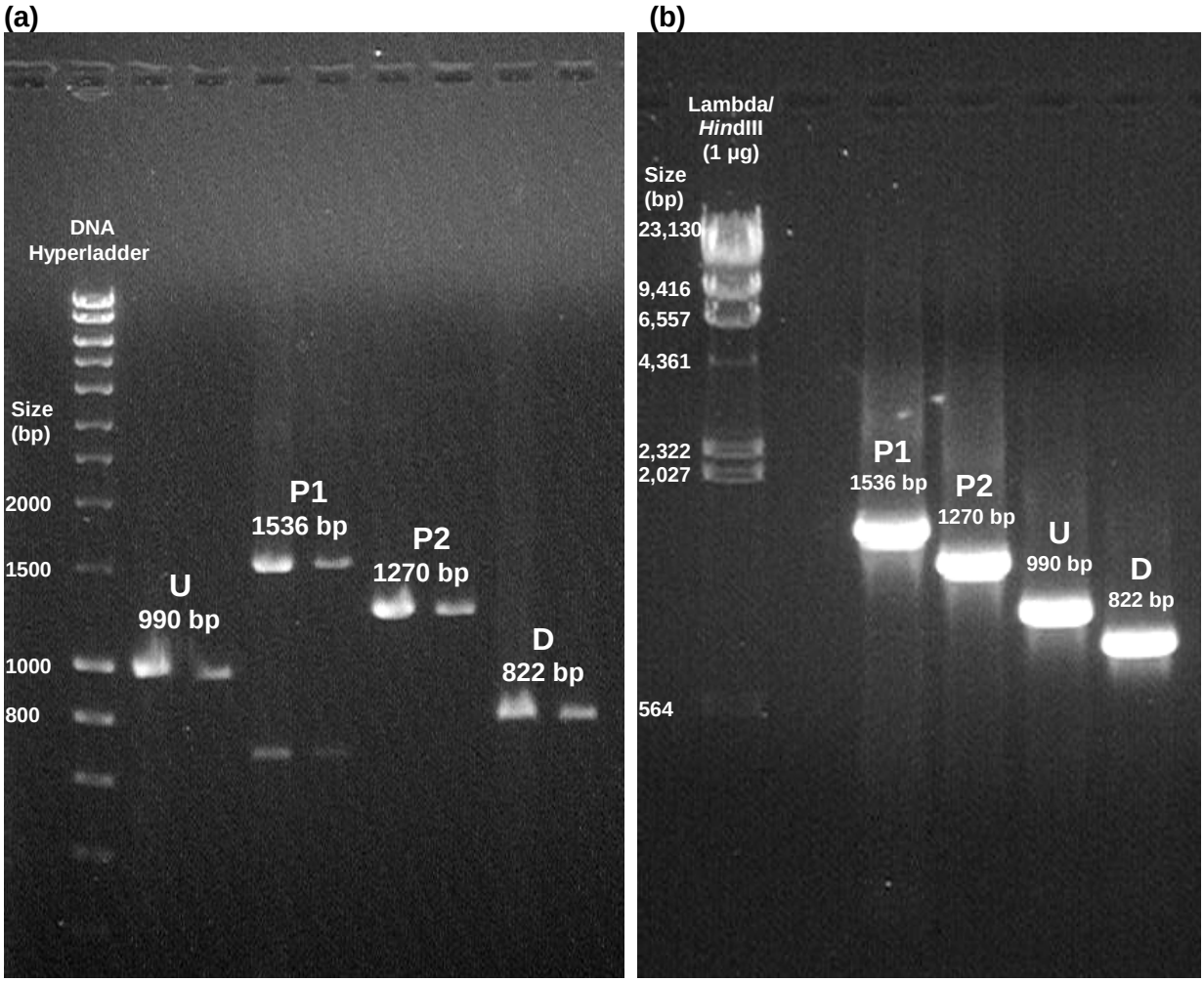


Figure 4.8. PCR-amplified *stzA*-flanking and *pyr4* gene fragments. DNA fragments upstream (U) and downstream (D) of the *stzA* gene were amplified by PCR using *A. nidulans* L20 genomic DNA as a template. Fragments of the *N. crassa pyr4* gene, designated P1 and P2, were also amplified by PCR using plasmid pDEL1 as a template. (a) Shows all four PCR-amplified fragments (10 µl and 5 µl aliquots) following agarose gel electrophoresis. (b) Shows these fragments (10 µl aliquots; ~1 µg DNA) after purification and concentration by gel band excision using a Qiagen DNA purification kit.

4.2.5 Generation of bipartite gene-targeting substrates by fusion PCR

Generation of the two *stzA*-gene targeting substrates was performed using two independent fusion PCR reactions. This procedure involved fusing fragment U with P1, and D with P2. Hence, *stzA*-flanking regions were fused with portions of the *pyr4* auxotrophic marker to generate two fusion products to be used to target the *stzA* gene. The corresponding nucleotide sequences are shown in Figures 4.9 and 4.12. A summary of the DNA fragment sizes is shown in Table 4.3.

Fusion PCR was attempted by two methods designated either “normal” or “nested”. Both types of fusion PCR involved the annealing of complementary fusion tags. Normal fusion PCR involved selecting the appropriate pairs of the original primers used to generate the initial PCR fragments. Nested fusion PCR involved the design of new primers within the fused sequences to be amplified. Both types of fusion PCR involved two thermocycler loops. The first loop involved using the annealing temperature at which complementary fusion tags would anneal (55 ºC for both fusion tags; Nielsen *et al*., 2006). The second loop involved optimising the annealing temperatures for primer pairs used to amplify the fusion products. These temperatures had to be optimised independently for the two fusion reactions along with other key PCR variables.

Table 4.3. Sizes of fusion products following fusion PCR. Note that the sizes of fusion products were shorter than the combined sizes of the initial fragments. This was due to the annealing of complementary fusion tags (one fusion tag sequence was lost from each pair following fusion). Further shortening of fusion products occurred as a result of nested PCR in which primers were selected within fused sequences rather than at their ends.

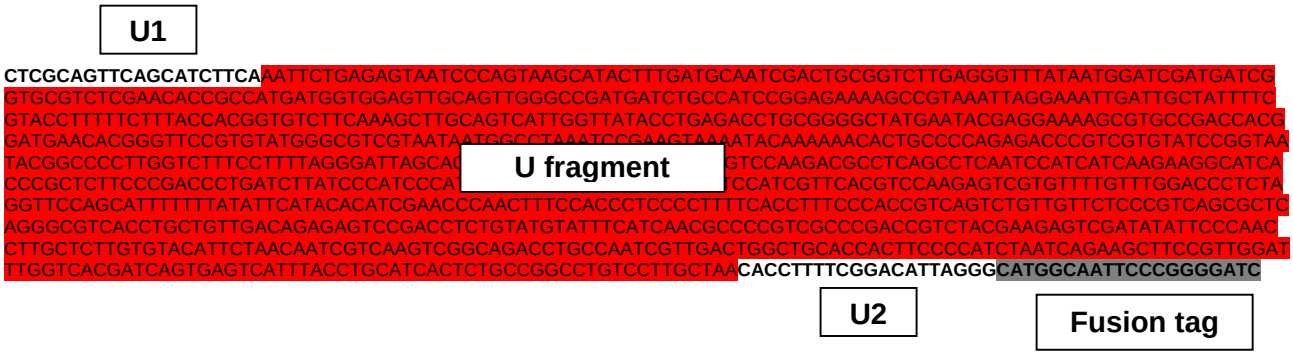
<table>
<tr><th rowspan="2">Fragment</th><th rowspan="2">Size (bp)</th><th colspan="2">Fusion product size (bp)</th></tr>
<tr><th>Normal fusion PCR</th><th>Nested fusion PCR</th></tr>
<tr><td>U</td><td>990</td><td rowspan="2">2508
(Primers U1 and M3)</td><td rowspan="2">2321
(Primers F1 and F2)</td></tr>
<tr><td>P1</td><td>1536</td></tr>
<tr><td>P2</td><td>1270</td><td rowspan="2">2074
(Primers M2 and D2)</td><td rowspan="2">1944
(Primers F3 and F4)</td></tr>
<tr><td>D</td><td>822</td></tr>
</table>

Attempts were made to fuse fragments U and P1 to yield fusion product 1, one of the bipartite substrates to be used for targeting *stzA*. Primers U1 and M3 were used for this purpose (Figure 4.9). No bands corresponding to a fusion product size of 2,508 bp were visible following changes to primer annealing temperature (54–60 ºC), DNA concentration (50–200 ng), $MgCl_2$ concentration (1.5 mM–3.5 mM) and fusion cycle number (2–15; Figure 4.10). High molecular weight DNA smears were apparent in each lane.

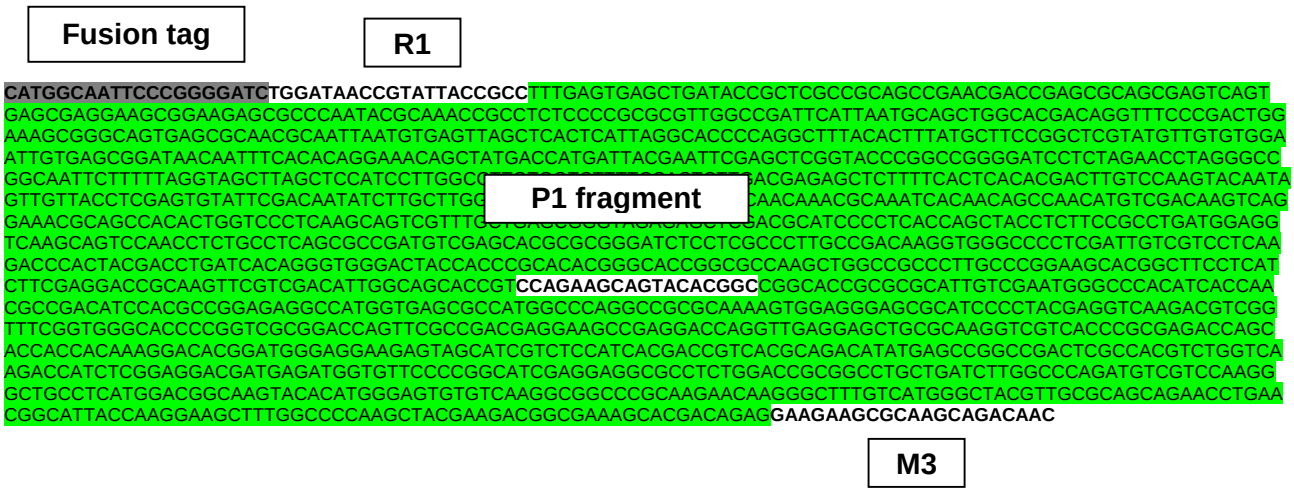
Fusion product 1, however, could be synthesised using primers F1 and F2 (nested fusion PCR) to yield a shorter product of 2,321 bp. High yields of this fusion product were obtained using a primer annealing temperature of 58 ºC, 10 fusion cycles and an $MgCl_2$ concentration of 2.5 mM (Figure 4.11).

Figure 4.9. Nucleotide sequence of fusion product 1. Primers are highlighted in white. Fusion tags are highlighted in grey. The U and P1 fragments are highlighted in red and green, respectively.

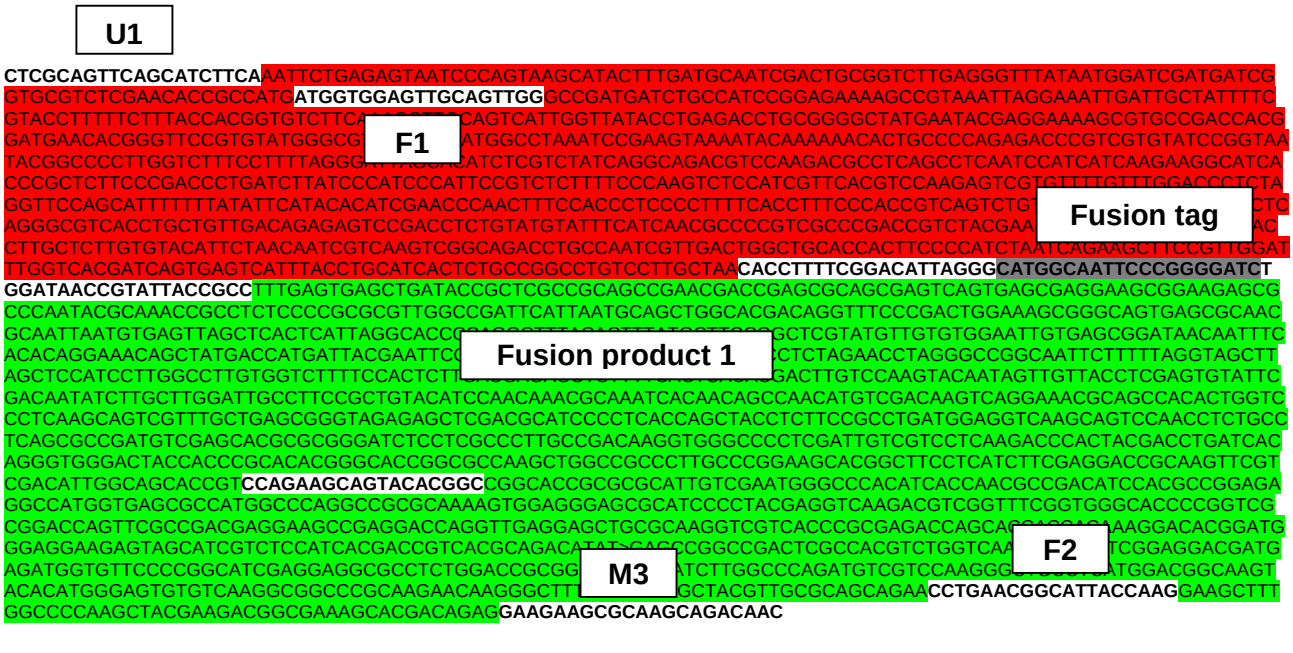
(a) Upstream region flanking the *stzA* gene. Amplification was performed using primers U1 and U2 to yield a 990 bp fragment (U).



(b) N-terminal ⅔ of the *pyr4* gene. Amplification was performed using primers R1 and M3 to yield a 1536 bp fragment (P1) containing the N-terminal ⅔ of the *pyr4* gene and a 292 bp direct repeat sequence.



(c) Fusion of U with P1. Fusion and amplification of U and P1 was attempted by fusion PCR using primer pairs U1 and M3 (normal PCR), and F1 and F2 (nested PCR) to yield expected fusion product sizes of 2,508 bp and 2,321 bp, respectively.



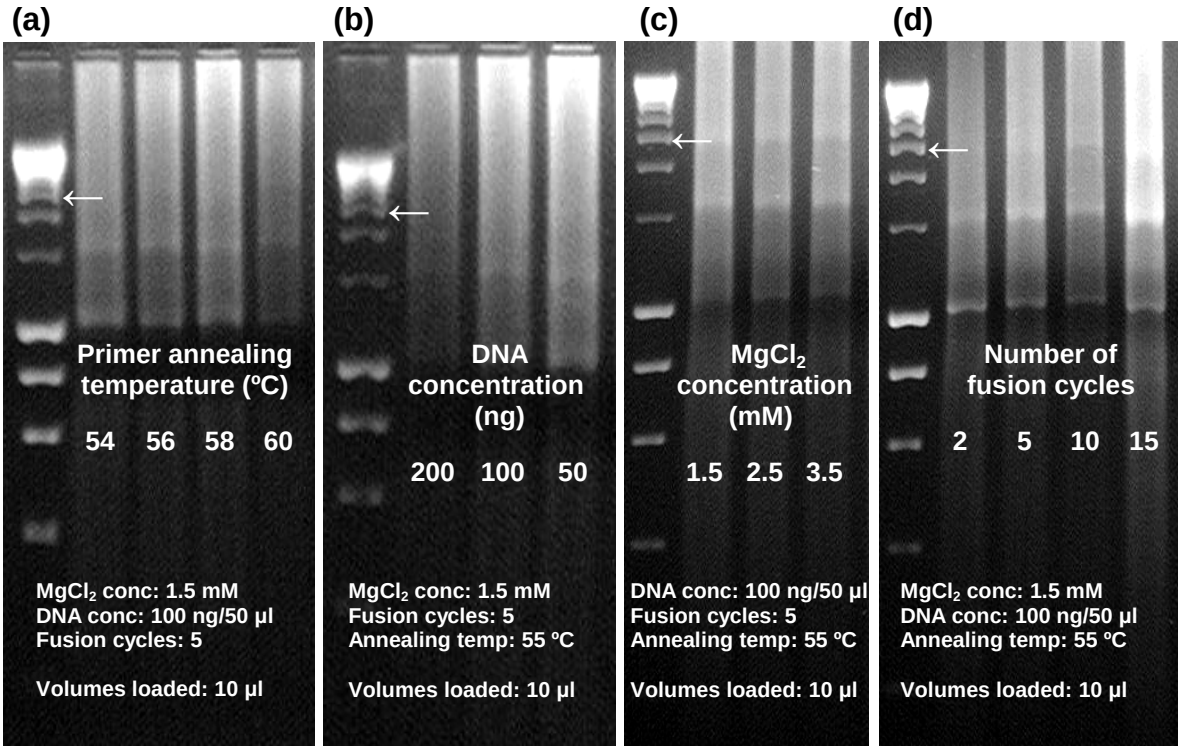


Figure 4.10. Attempts to synthesise fusion fragment 1 by normal fusion PCR. Synthesis of fusion fragment 1 was attempted using primers U1 and M3 by changes to (a) primer annealing temperature, (b) DNA concentration, (c) $MgCl_2$ concentration and (d) fusion cycle number. Note that where DNA concentrations refer to 100 ng / 50 µl, this means 100 ng of each fragment. A white arrow indicates the 2.5 kb DNA band of the hyperladder, a region where the 2,508 bp fusion product was expected to be visible.

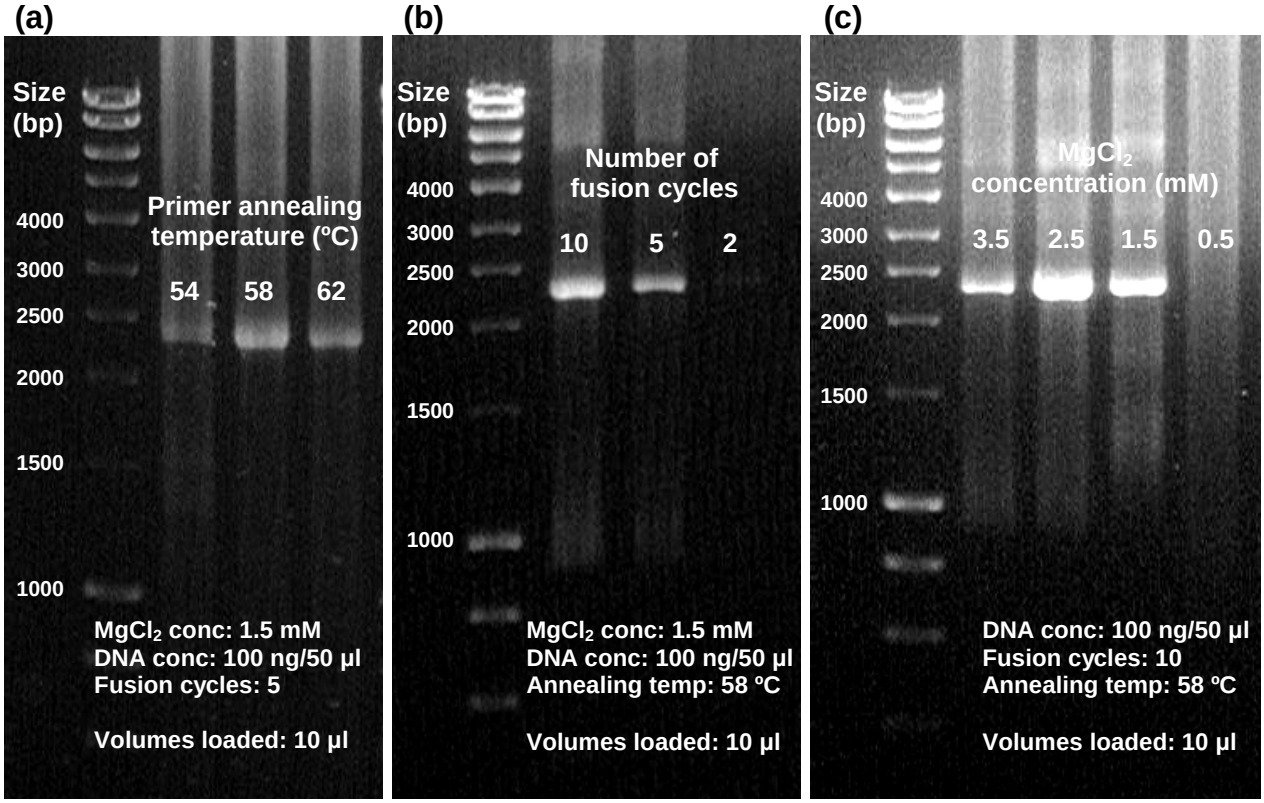


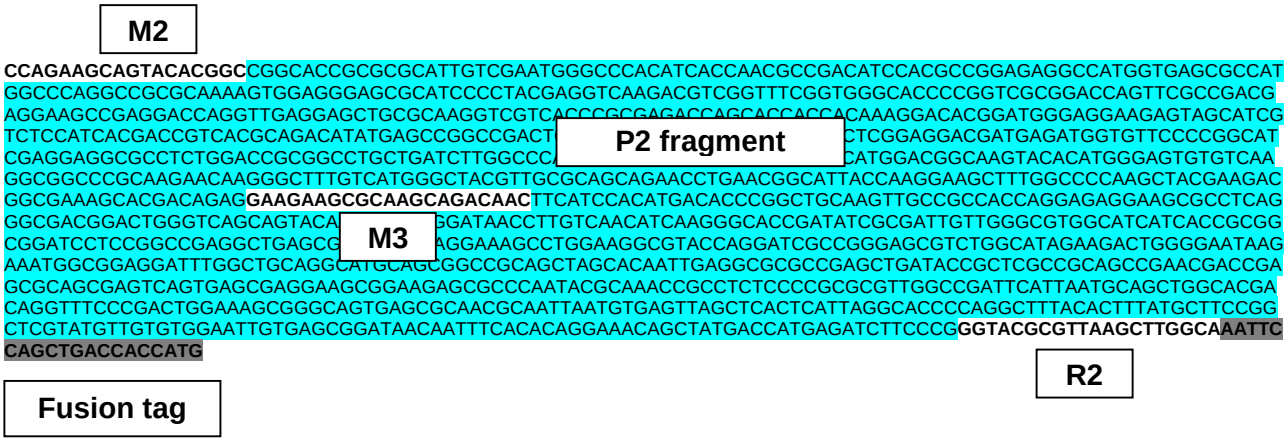
Figure 4.11. Optimisation of fusion fragment 1 synthesis by nested fusion PCR. Primers F1 and F2 were used to perform nested fusion PCR to yield fusion fragment 1 with a size of 2,321 bp. Optimisation of fusion product formation involved changes to (a) primer annealing temperature, (b) fusion cycle number and (c) $MgCl_2$ concentration.

Attempts were then made to fuse fragments P2 and D to yield fusion product 2, the second bipartite substrate to be used for targeting *stzA*. Primers M2 and D2 were used for this purpose (Figure 4.12). The optimum annealing temperature, $MgCl_2$ concentration and DNA concentration were determined to be 54 ºC, 2.5 mM and 50 ng / 50 µl, respectively. A summary of the successful aspects of this optimisation procedure is shown by Figure 4.13, which nevertheless shows that only low yields of fusion product 2 (2,074 bp) were obtained.

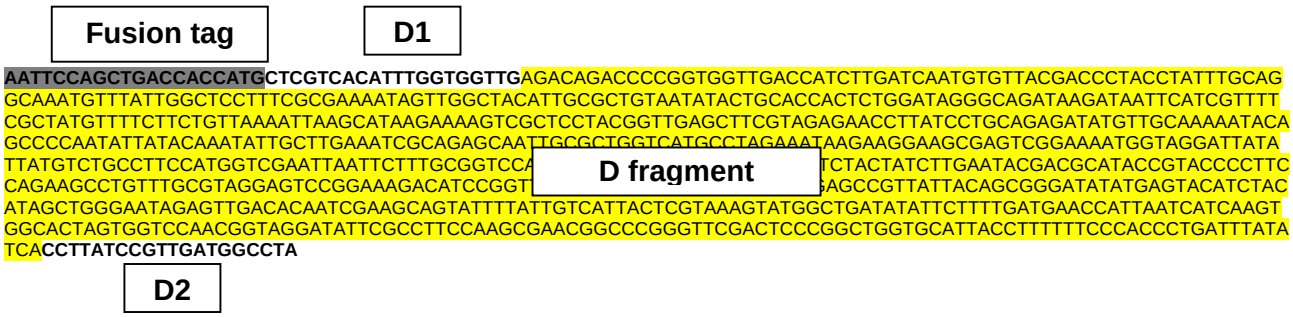
Greatly improved yields of fusion product 2 (1,944 bp) could be obtained using a nested fusion PCR approach involving primers F3 and F4 under optimised PCR conditions (an annealing temperature of 54 ºC, a DNA concentration of 100 ng / 50 µl and an $MgCl_2$ concentration of 1.5 mM; Figure 4.14).

Figure 4.12. Nucleotide sequences of fusion product 2. Primers are highlighted in white. Fusion tags are highlighted in grey. The P2 and D fragments are highlighted in blue and yellow, respectively.

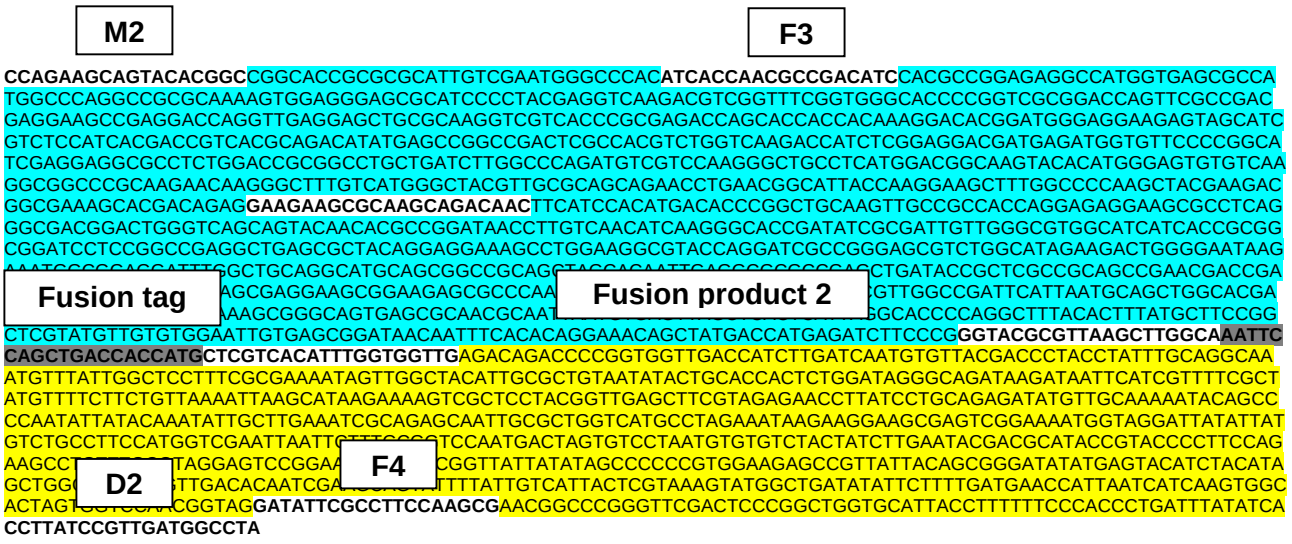
(a) C-terminal ⅔ of the *pyr4* gene. Amplification was performed using primers M2 and R2 to yield a 1270 bp fragment (P2) containing the C-terminal ⅔ of the *pyr4* gene and a 292 bp direct repeat sequence.



(b) Downstream region flanking the *stzA* gene. Amplification was performed using primers D1 and D2 to yield an 822 bp fragment (D).



(c) Fusion of P2 with D. Fusion and amplification of P2 and D was attempted by fusion PCR using primer pairs M2 and D2 (normal PCR), and F3 and F4 (nested PCR) to yield expected fusion product sizes of 2,074 bp and 1,944 bp, respectively.



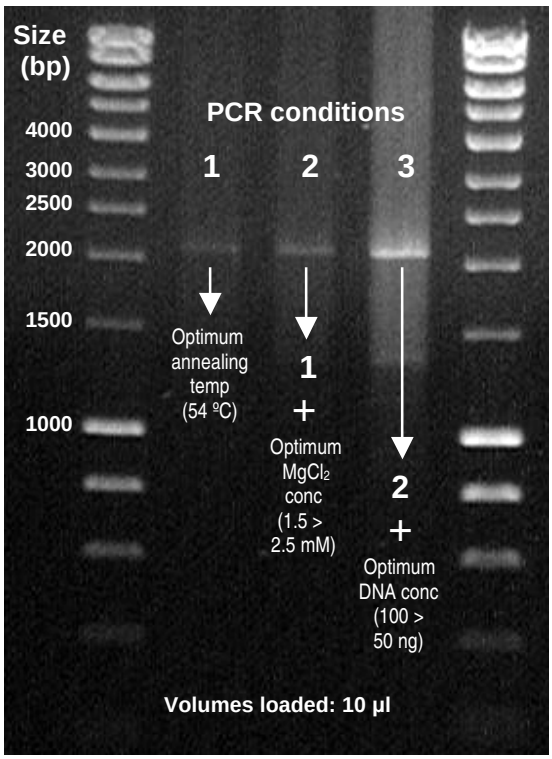


Figure 4.13. Synthesis of fusion fragment 2 by normal fusion PCR. Synthesis of fusion fragment 2 was performed using primers M2 and D2 to yield a 2,074 bp fusion product. The optimum annealing temperature, $MgCl_2$ concentration and DNA concentration were determined to be 54 ºC, 2.5 mM and 50 ng / 50 µl, respectively. The minimal increase of fusion product formation as a result of combining these optimal conditions is shown (conditions 1, 2 and 3). An increase in the number of fusion cycles from 5 to 10 was also attempted (not shown) but this did not increase fusion product formation.

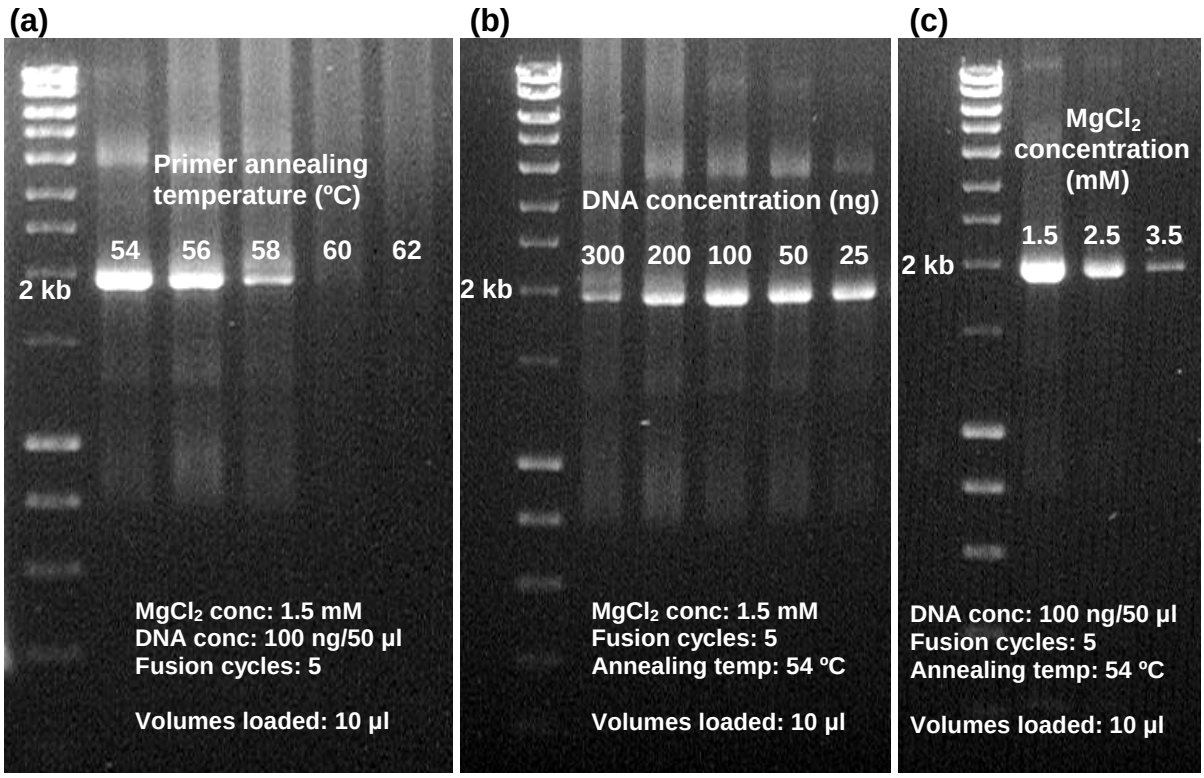


Figure 4.14. Optimisation of fusion fragment 2 synthesis by nested fusion PCR. Use of the nested primers F3 and F4 enhanced fusion product formation (1,944 bp) compared with primers M2 and D2. Maximum product formation was achieved by optimising (a) annealing temperature, (b) DNA concentration and (c) $MgCl_2$ concentration.

Optimised fusion PCR reaction and thermocycler conditions for the synthesis of both fusion products 1 and 2 using nested PCR are shown in Table 4.4. The DNA bands were subsequently run along a gel, excised and purified using a Qiagen DNA purification kit to remove primers and any contaminating DNA fragments. The purified fusion products were again run on a gel for quantification prior to transformation (Figure 4.15). Both DNA fragments were present at concentrations of ~20 ng μl^{-1}.

Table 4.4. Optimised fusion PCR conditions. The PCR conditions used to fuse U with P1 and D with P2 are shown below. The first thermocycler loop involved 5–10 long-duration (5 mins) fusion cycles to enable the fusion of complementary fusion tags at 55 ºC for both pairs. This was followed by 40 cycles using the optimum annealing temperatures for fusion product amplification (58 and 54 ºC for fusion products 1 and 2, respectively. The 40 cycles were split into two loops of 20 cycles for fusion product 1 so that extension times could be increased from 4 mins 30 sec to 6 mins, which enabled more product to be formed. During the optimisation of fusion product formation, 25 µl reaction volumes were used in PCR reactions and 10 µl aliquots were run on a gel with 2 µl of 6x loading buffer. During the purification process, total reaction volumes of 200 µl to 400 µl (25 µl aliquots) were run along multiple lanes of a gel and purified by band excision using a Qiagen DNA purification kit.

(a) Reaction conditions

Component	**Stock conc**	**Final conc**	**Fusion reaction**	
			1	**2**
SDW			124.4 µl	144.4 µl
5x HF buffer + $MgCl_2$	5x 7.5 mM $MgCl_2$	1x 1.5 mM $MgCl_2$	40 µl	40 µl
$MgCl_2$	10 mM	1.5 OR 2.5 mM	20 µl for nested PCR	No extra $MgCl_2$
dNTPs	25 mM	0.2 mM	1.6 µl	1.6 µl
Primers	50 µM	0.5 µM	U1 and M3 – 2 µl each OR F1 or F2 – 2 µl each	M2 and D2 – 2 µl each OR F3 and F4 – 2 µl each
Template DNA	100 ng μl^{-1}	2 ng μl^{-1}	U fragment – 4 µl P1 fragment – 4 µl	D fragment – 4 µl P2 fragment – 4 µl
Phusion® polymerase	2 U μl^{-1}	0.01 U μl^{-1}	2 µl	2 µl
Total volume			= 200 µl	= 200 µl

(b) Thermocycler conditions

Fusion reaction	
1	**2**
Initial denaturing 94 ºC – 5 mins **Loop 1 – Fusion** 10 cycles 94 ºC – 30 sec 55 ºC – 5 min 72 ºC – 3 min 30 sec **Loop 2 – Primer annealing** 20 cycles 94 ºC – 30 sec 58 ºC – 30 sec 72 ºC – 4 mins 30 sec **Loop 3 – Primer annealing** 20 cycles 94 ºC – 30 sec 58 ºC – 30 sec 72 ºC – 6 min **Final extension** 72 ºC – 15 min **Final hold (4 ºC)**	**Initial denaturing** 94 ºC – 5 min **Loop 1 – Fusion** 5 cycles 94 ºC – 30 sec 55 ºC – 5 min 72 ºC – 3 min **Loop 2 – Primer annealing** 40 cycles 94 ºC – 30 sec 54 ºC – 30 sec 72 ºC – 4 min **Final extension** 72 ºC – 15 min **Final hold (4 ºC)**

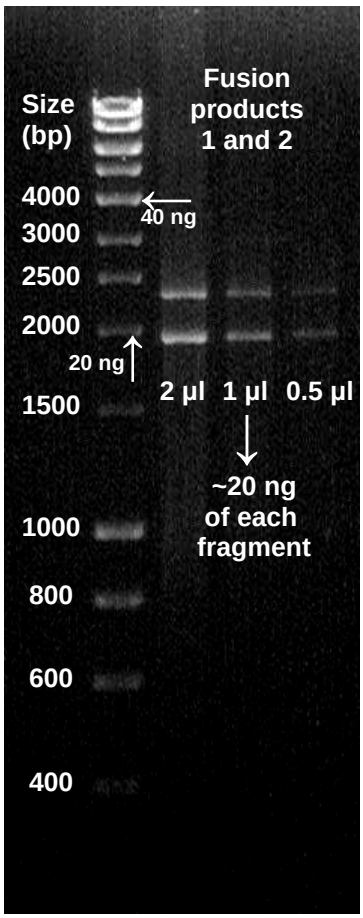


Figure 4.15. Purified fusion products 1 and 2. These fusion products (2,321 and 1,944 bp) were the bipartite gene-targeting substrates used for targeting and deleting the *stzA* gene. Approximately 10 µl of each fragment (~200 ng) was used to transform 2 x 10^7 *A. nidulans* protoplasts from both A1149 and G191.

4.2.6 Verification of correct bipartite gene-targeting substrates

Restriction digests were performed to verify that the correct DNA sequences had been amplified and fused. Each fusion product was cut individually with four restriction enzymes in separate digests. The expected sizes of the restricted fragments were calculated at NEBcutter (http://tools.neb.com/NEBcutter2/) and are shown above or below each fragment in Figure 4.16. All prominent restriction fragments were of the expected sizes. Note that some non-specific products were present in high weight molecular smears, which may have undergone restriction because unpurified fusion products (readily available in larger quantities) were used for restriction purposes.

(a) Fusion product 1 **(b) Fusion product 2**

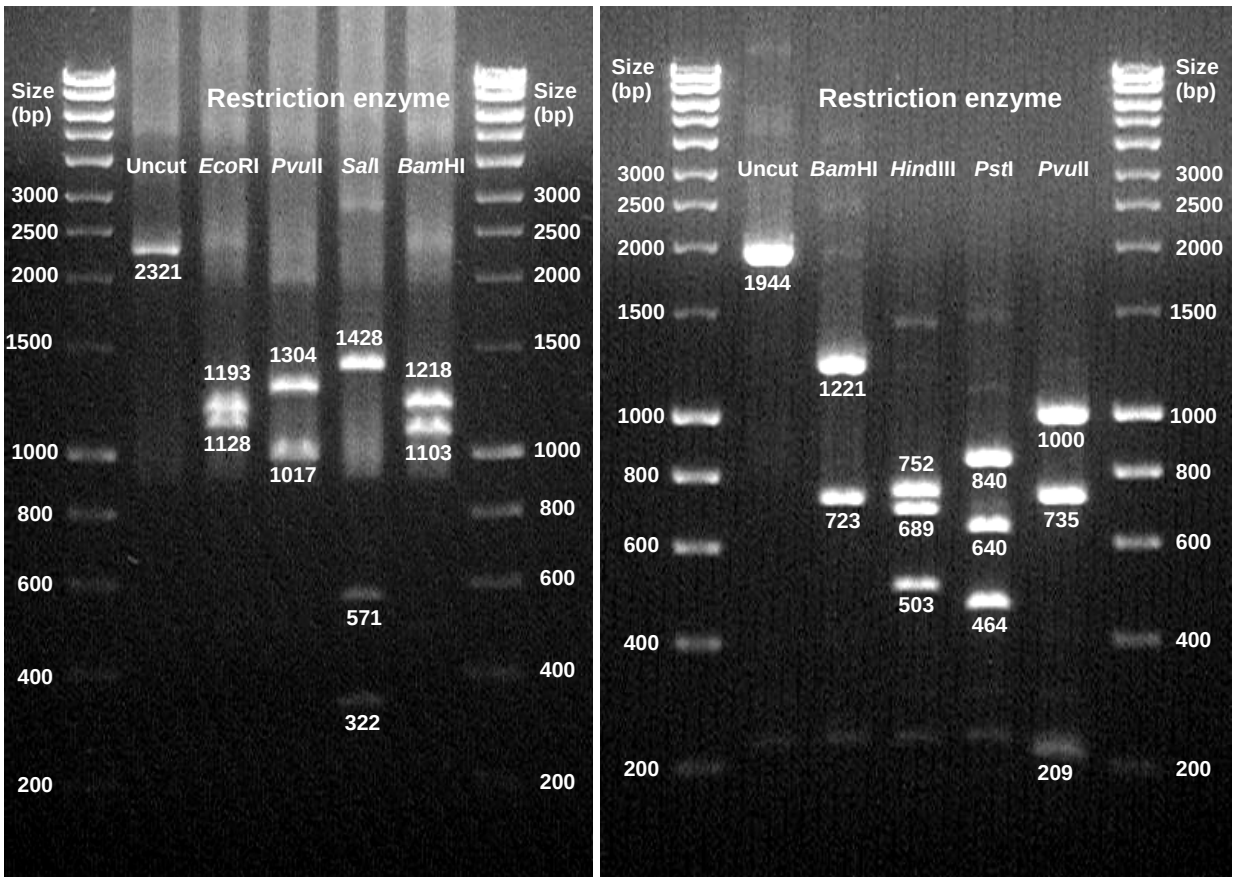


Figure 4.16. Restriction digests of fusion products 1 and 2. Four single-enzyme digests were performed for each fusion product. Approximately 200–300 ng of each restricted product is shown for each digest.

4.2.7 Transformation of G191 and A1149 with pDJB3 and *stzA* gene-targeting substrates

Transformation of *A. nidulans* G191 and A1149 (~2 x 10^7 protoplasts) was optimised using pDJB3 (~1 µg) before targeting the *stzA* gene of these strains with the two bipartite fusion substrates. pDJB3 contains the *pyr4* gene from *N. crassa* and complements the mutated *pyrG* gene in *pyrG89* strains, such as G191 and A1149, allowing transformants to grow on regeneration medium lacking uridine and uracil.

Initial transformations, based on the method of Tilburn *et al*. (1983), resulted in only 9–16 transformants per µg of plasmid DNA for G191 protoplasts in three similar experiments. Optimisation of the transformation procedure involved reducing the agar concentration in the regeneration overlay medium from 1.5 to 0.5%, plating protoplasts directly onto regeneration medium, and adding 0.6 M KCl to the PEG solution (25% PEG 6000, 50 mM $CaCl_2$, 10 mM Tris, pH 7.5). By combining direct plating with the use of osmotically stabilised PEG solution, approximately 250 transformants / µg of plasmid were routinely obtained for G191, and similar numbers were subsequently obtained for A1149 (Figure 4.17).

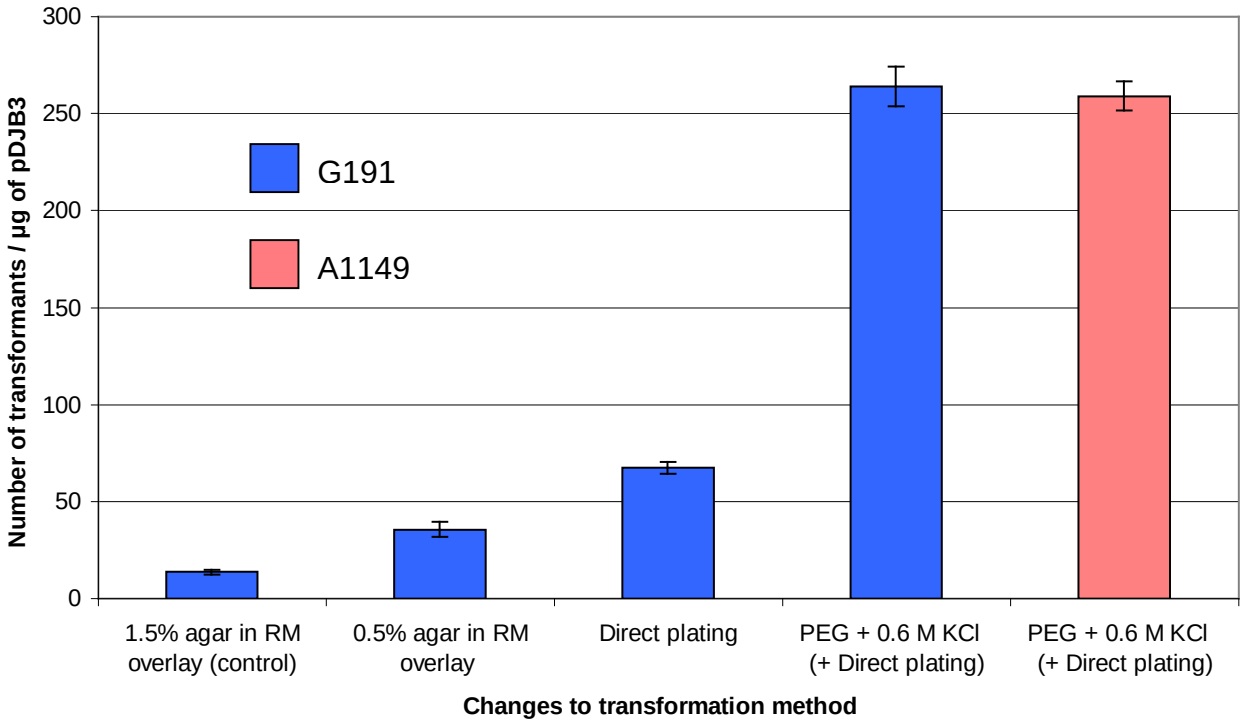


Figure 4.17. Optimisation of the transformation efficiency of *A. nidulans* G191 with pDJB3. Numbers of transformants are the means of triplicate experiments in which 2 x 10^7 protoplasts were transformed with 1 µg of pDJB3. Error bars indicate standard errors. RM = regeneration medium.

Transformation of A1149 and G191 with the gene-targeting fusion substrates yielded 11 and 28 transformants for each strain, respectively. These were numbered A1–A11 and G1–G28. All 39 strains were re-plated onto selective medium containing 0.6 M KCl. Only a total of five of the strains regenerated; the remainder were abortive transformants. The regenerated strains A2, A5, A7, G6 and G25 were assumed to be *stzA* deletants and were therefore analysed phenotypically. Two stable pDJB3 transformants for both A1149 (designated AP1 and AP2) and G191 (GP1 and GP2) were set aside to be used as control strains.

4.2.8 PCR assay confirming *stzA* deletion from transformants

Assuming that the *stzA* gene had been successfully deleted from each of the five transformants, the sequence surrounding this locus (now *pyr4*) was deduced (Figure 4.18). A PCR assay was then designed to verify that the *stzA* ORF had been deleted from each of the transformants. Figure 4.19 shows the locus and DNA sequence of the PCR assay product (2,437 bp; Figure 4.19) amplified using primers S1 and S2 from the extracted genomic DNA of the transformants (Figure 4.20). The PCR conditions used to amplify the PCR product are shown in Table 4.5. As the PCR product could be amplified from all five *stzA* deletion transformants (Figure 4.21), it seems likely that successful *stzA* deletion occurred in all strains.

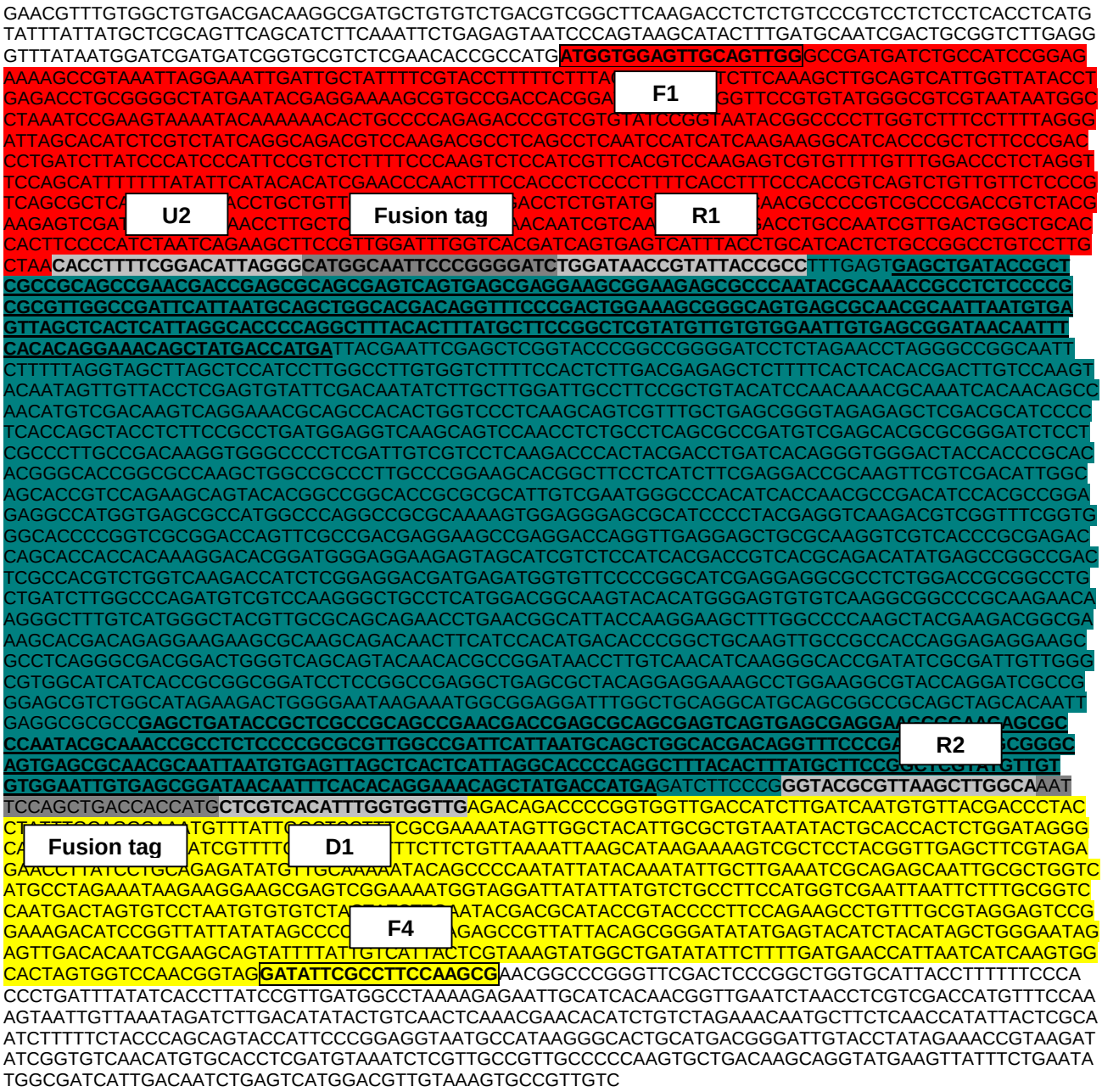
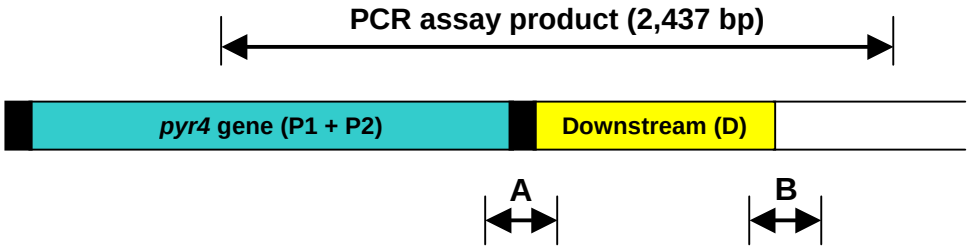


Figure 4.18. Genomic sequence surrounding the *pyr4* locus in *stzA* deletion transformants. In *stzA* deletion strains, the entire *stzA* ORF (2,202 bp) and 11 additional bp were replaced with the *pyr4* gene as a result of three HR events. Two of these events occurred between the genomic DNA and the U (red) and D (yellow) portions of the bipartite gene targeting substrates (fusion products). The *pyr4* gene (turquoise), which replaced the *stzA* ORF, was complete and functional following HR of the overlapping ⅔ portions of the *pyr4* gene contained within P1 and P2. The two 292 bp direct repeat sequences that flank the *pyr4* gene are shown bold and underlined. Sequences shown that are not highlighted (white) comprise the original genomic DNA sequences. Primer sequences and fusion tags are highlighted in light and dark grey, respectively.

(a)



(b)

CCAGAAGCAGTACACGGCCGGCACCGCGCGCATTGTCGAATGGGCCCACATCACCAACGCCGACATCCACGCCGG
AGAGGC S1 AGCGCCATGGCCCAGGCCGCGCAAAAGTGGAGGGAGCGCATCCCCTACGAGGTCAAGACGT
CGGTTT CACCCCGGTCGCGGACCAGTTCGCCGACGAGGAAGCCGAGGACCAGGTTGAGGAGCTGCGC
AAGGTCGTCACCCGCGAGACCAGCACCACCACAAAGGACACGGATGGGAGGAAGAGTAGCATCGTCTCCATCACG
ACCGTCACGCAGACATATGAGCCGGCCGACTCGCCACGTCTGGTCAAGACCATCTCGGAGGACGATGAGATGGTG
TTCCCCGGCATCGAGGAGGCGCCTCTGGACCGCGGCCTGCTGATCTTGGCCCAGATGTCGTCCAAGGGCTGCCTC
ATGGACGGCAAGTACACATGGGAGTGTGTCAAGGCGGCCCGCAAGAACAAGGGCTTTGTCATGGGCTACGTTGCG
CAGCAGAACCTGAACGGCATTACCAAGGAAGCTTTGGCCCCAAGCTACGAAGACGGCGAAAGCACGACAGAGGAA
GAAGCGCAAGCAGACAACTTCATCCACATGACACCCGGCTGCAAGTTGCCGCCACCAGGAGAGGAAGCGCCTCAG
GGCGACGGACTGGGTCAGCAGTACAACACGCCGGATAACCTTGTCAACATCAAGGGCACCGATATCGCGATTGTTG
GGCGTGGCATCATCACCGCGGCGGATCCTCCGGCCGAGGCTGAGCGCTACAGGAGGAAAGCCTGGAAGGCGTAC
CAGGATCGCCGGGAGCGTCTGGCATAGAAGACTGGGGAATAAGAAATGGCGGAGGATTTG**GCTGCAGGCATGCA**
GCGGCCGCAGCTAGCACAATTGAGGCGCGCCGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCG
AGTCAGTGAGCGAGGAAGCGGAAGAGCGCC REGION A **CCTCTCCCCGCGCGTTGGCCGATTCATTAA**
TGCAGCTGGCACGACAGGTTTCCCGACTGGAGCGCAACGCAATTAATGTGAGTTAGCTCAC
TCATTA D1 **CAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTT**
CACACAGCTATGACCATGAGATCTTCCCGGGTACGCGTTAAGCTTGGCAAATTCCAGCTGACCACCAT
GCTCGTCACATTTGGTGGTTGAGACAGACCCCGGTGGTTGACC R2 TCAATGTGTTAC Fusion tag TTT
GCAGGCAAATGTTTATTGGCTCCTTTCGCGAAAATAGTTGGCTGCTGTAATATAC**GGA**
TAGGGCAGATAAGATAATTCATCGTTTTCGCTATGTTTTCTTCTGTTAAAATTAAGCATAAGAAAAGTCGCTCCTACG
GTTGAGCTTCGTAGAGAACCTTATCCTGCAGAGATATGTTGCAAAAATACAGCCCCAATATTATACAAATATTGCTTG
AAATC SA CAATTGCGCTGGTCATGCCTAGAAATAAGAAGGAAGCGAGTCGGAAAATGGTAGGATTATATTATG
TCTGC GGTC**GAATTAATTCTTTGCGGTCCAATGACTAGTGTCCTAATGTGTGTCTACTATCTTGAATACG**
ACGCATACCGTACCCCTTCCAGAAGCCTGTT REGION B **GAAAGACATCCGGTTATTATATAGCCCCCC**
GTGGAAGAGCCGTTATTACAGCGGGATATATAGCTGGGAATAGAGTTGACACAATCGAAGC
AGTATTTTATTGTCATTACTCGTAAAGTATGGCTGATATATTCTTTTGATGAACCATTAATCATCAAGTCGCACTAG
TGGTCCAACGGTAGGATATTCGCCTTCCAAGCGAACGGCCCGGGTTCGACTCCCGGCTGGTG SB TTTTTT
CCCACCCTGATTTATATCACCTTATCCGTTGATGGCCTAAAAGAGAATTGCATCACAACGGTTGAATCTAACCTCG
TCGACCATGTTTCCAAAGTA F4 ATAGATCTTGACATATACTGTCAACTCAAACGAACACATCTGTCTAGAAA
CAATGC S2 TGTACCTATAGAAACCGTAAGATATCGGTGTCAACATGTGCACCTCGATGTAAATCTCGTTGCC
GTTGCCCCCAAGTGCTGACAAGCAGGTATGAAGTTATTTCTGAATATGGCGATCATTGACAATCTGAGTCATGGACG
TTGTAAAGTGCCGTTGTC

Figure 4.19. Locus of the PCR assay product. (a) Shows the locus from which the PCR assay product (2,437 bp) was amplified from all five *stzA* transformants (A2, A5, A7, G6 and G25) using primers S1 and S2. Regions A and B of the PCR product were sequenced by the Dundee Sequencing Centre (http://www.dnaseq.co.uk) using primers SA and SB for the *stzA* deletion transformant A2. (b) Shows the nucleotide sequence of the PCR assay product.

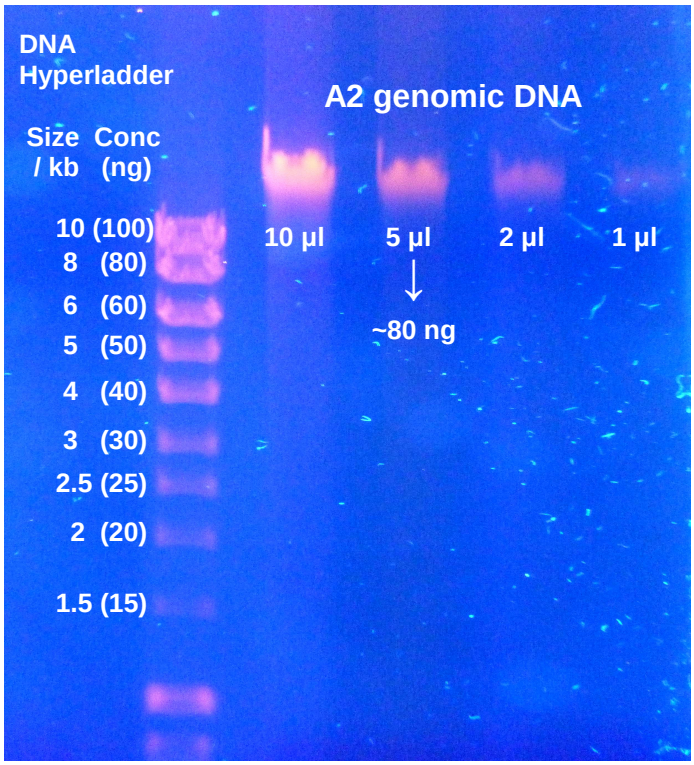


Figure 4.20. Quantification of genomic DNA from transformant A2. Genomic DNA extracted from all five potential *stzA* deletion transformants and untransformed A1149 was quantified in a similar manner to the example shown by comparison to known quantities of DNA bands in a hyperladder.

Table 4.5. PCR assay conditions. Reaction and thermocycler conditions were based on the guidelines in the Abgene® *Taq* DNA polymerase handbook.

(a) Reaction conditions

Constituents	Stock conc	Final conc	Volume
SDW			130.4 µl
PCR buffer	10x	1x	20 µl
$MgCl_2$	7.5 mM	1.5 mM	40 µl
dNTPs	25 mM	0.2 mM	1.6 µl
Primer S1	50 µM	0.5 µM	2 µl
Primer S2	50 µM	0.5 µM	2 µl
Genomic DNA	200 ng μl^{-1}	2 ng μl^{-1}	2 µl
***Taq* DNA polymerase**	5 U μl^{-1}	0.05 U μl^{-1}	2 µl
Total volume			= 200 µl

(b) Thermocycler conditions

Initial denaturing
94 ºC – 5 min

Loop 1
40 cycles
94 ºC – 30 sec
60 or 64 ºC – 30 sec
72 ºC – 4 min

Final extension
72 ºC – 15 mins

Final hold (4 ºC)

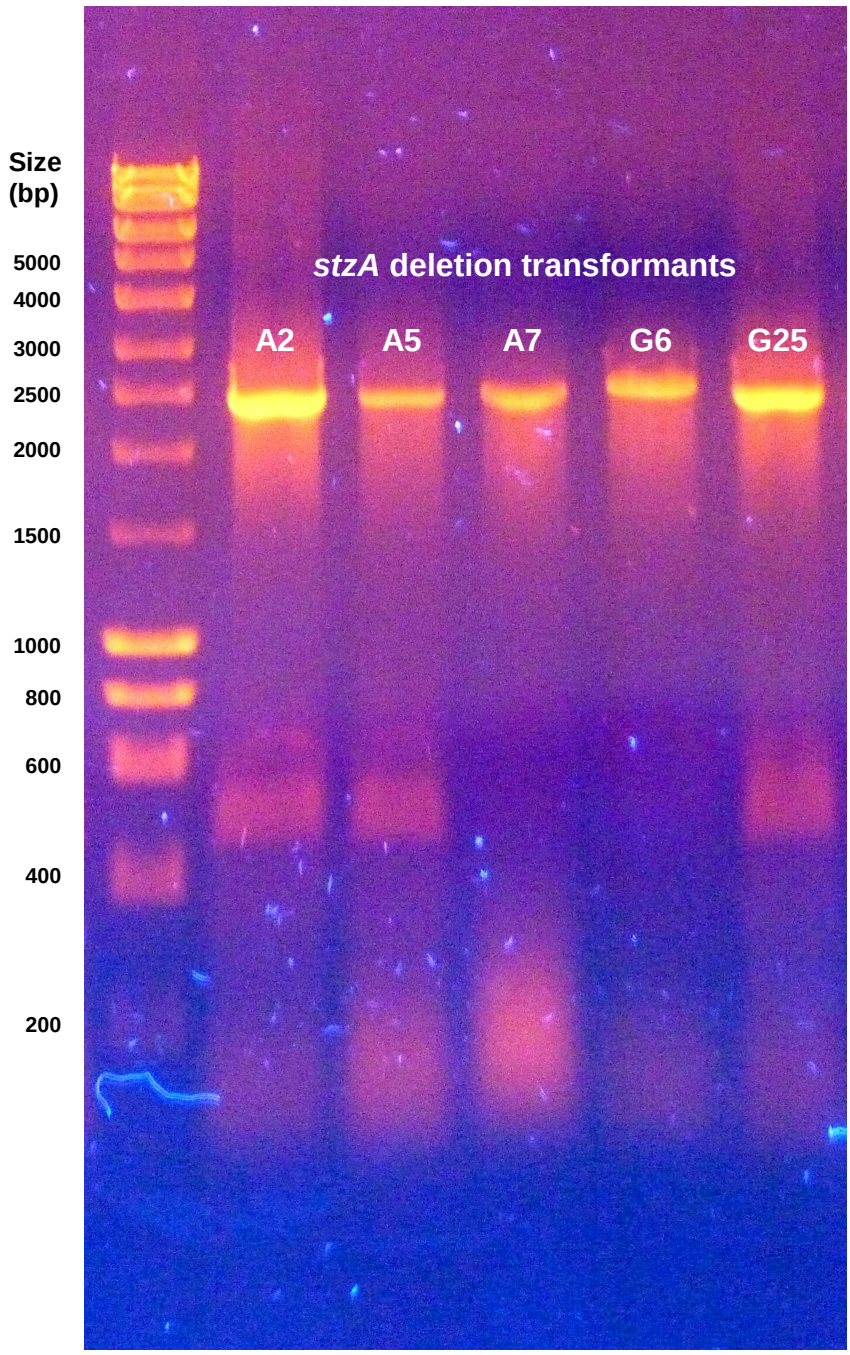


Figure 4.21. PCR assay product from *stzA* deletion transformants. The PCR assay product (2,437 bp) was successfully amplified from all five *stzA* deletion transformants (A2, A5, A7, G6 and G25) using primers S1 and S2. The assay helped to confirm that the *stzA* ORF had been deleted from each of the transformants, with correct fusion between the P2 and D fragments and integration of *pyr4* at the *stzA* locus.

Amplification of the intended PCR assay product was verified by the use of a control and by restriction analysis. Absence of the PCR assay product in the control strain A1149 (untransformed) verified that this product was present only in *stzA* deletion strains that included A2 (Figure 4.22a). Note that the annealing temperature was increased from 60 to 64 ºC (for both A2 and A1149). This increase resulted in the absence of non-specific PCR products in the A2 reaction, avoiding the need for DNA band excision from the gel and subsequent purification prior to restriction digests. The A2 PCR assay product was digested with *Bam*HI (single digest; b) and both *Bam*HI and *Xba*I (double digest; c) The expected sizes of the restricted fragments were calculated at NEBcutter (http://tools.neb.com/NEBcutter2/) and are shown above or below each fragment. All restriction fragments were of the expected sizes.

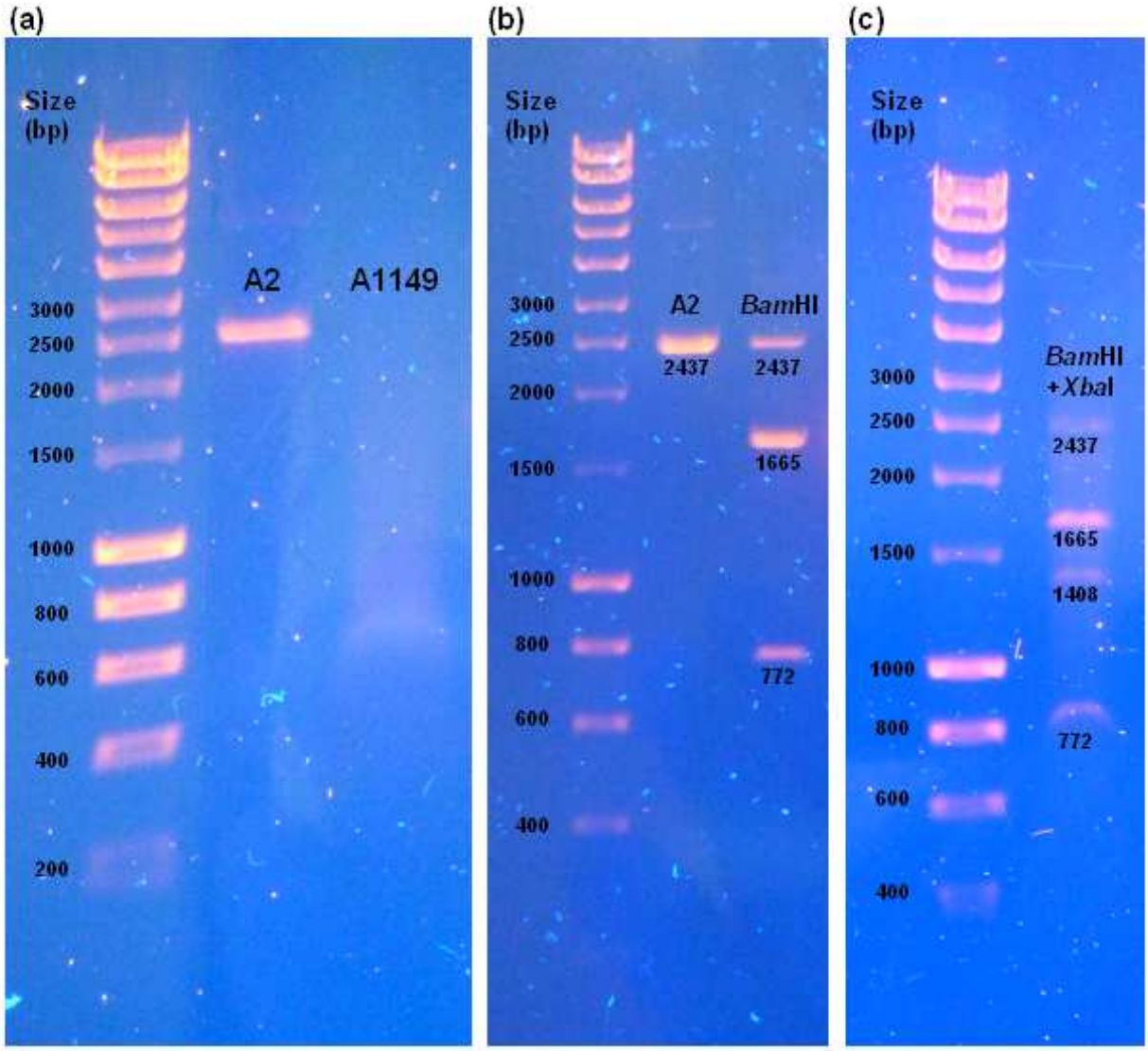


Figure 4.22. Verification of PCR assay product using a control and restriction analysis. (a) The PCR assay product for the *stzA* deletion transformant A2 and the control strain A1149 using an annealing temperature of 64 ºC. (b) The PCR assay product restricted with *Bam*HI and (c) *Bam*HI and *Xba*I.

Regions A and B of the PCR assay product (Figure 4.19) were sequenced by the Dundee Sequencing Centre (http://www.dnaseq.co.uk) for the *stzA* deletion transformant A2. Primers SA and SB were designed and used by the Dundee Sequencing Centre to sequence both regions by direct sequencing. Sequence data for these regions were aligned with the expected data using ClustalW (Figures 4.23 and 4.24). Sequence data for region A (558 nucleotides) indicated that the P2 and D fragments had fused correctly to yield one of the bipartite gene-targeting substrates. Furthermore, this region was sequenced with 100% accuracy. Data for region B indicated that the *pyr4* gene had correctly integrated at the *stzA* locus to knockout the *stzA* ORF. This region was sequenced with 99.6% accuracy. The chromatograms showing the sequence data for regions A and B are shown in Appendix 5.

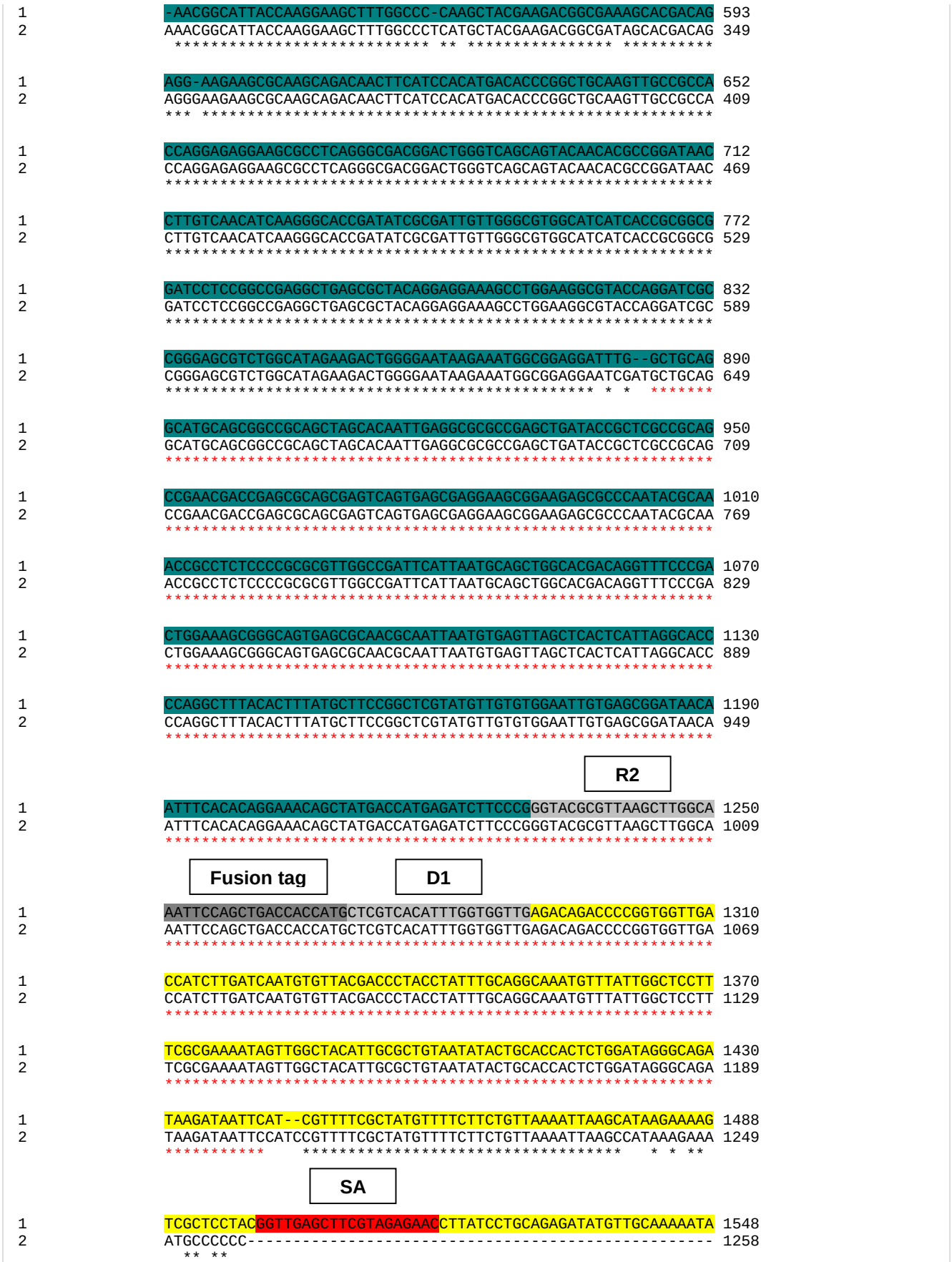


Figure 4.23. ClustalW alignment showing (1) the expected sequence and (2) sequence data for region A. Sequence data obtained for nucleotides 643–1200 (558 nucleotides) is 100% accurate compared to the expected sequence (red asterisks). The reverse complement (3'-GGTTGAGCTTCGTAGAGAAC) of primer SA (5'-GTTCTCTACGAAGCTCAACC) designed and used by the Dundee Sequencing Centre to sequence this region is highlighted in red. Portions of the *pyr4* gene (green) and fragment D (yellow) are highlighted on the expected sequence. The fusion tag and primers R2 and D1 are highlighted in dark and light grey.

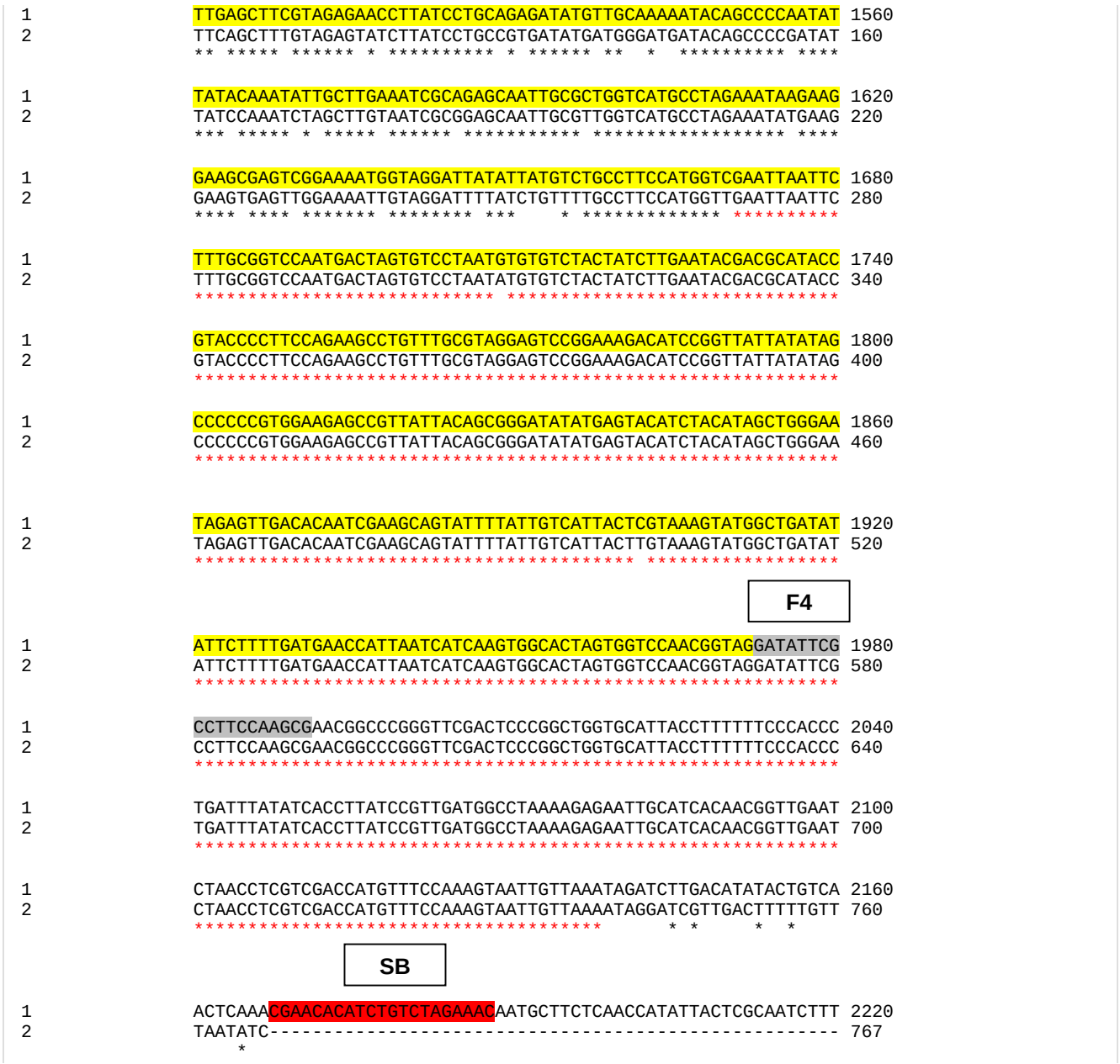


Figure 4.24. ClustalW alignment showing (1) the expected sequence and (2) sequence data for region B. Sequence data obtained for nucleotides 271–738 (468 nucleotides) is 99.6% accurate compared to the expected sequence (red asterisks). The reverse complement (3'-CGAACACATCTGTCTAGAAAC-5') of primer SB (5'- GTTTCTAGACAGATGTGTTCG -3') designed and used by the Dundee Sequencing Centre to sequence this region is highlighted in red. A portion of fragment D (yellow) is highlighted on the expected sequence. *A. nidulans* genomic DNA sequence is shown in white. Primer F4 is highlighted in light grey.

4.3 Results: Characterisation of *stzA* deletion strains

4.3.1 Osmotic stress responses

Growth of *A. nidulans* strains on media imposing ionic osmotic stresses was analysed. The *stzA* deletion mutants (A2, A5, A7, G6 and G25) and the *sltA1* mutant GO281 were more sensitive to the osmotic stresses exerted by sodium chloride (0.5 M), potassium chloride (0.6 M), magnesium chloride (0.2 M), lithium chloride (0.3 M), aluminium chloride (0.15 M) and rubidium chloride than the control strains (AP1, AP2, GP1 and GP2) and the wild-type strain L20 (Figure 4.25). Addition of 1% of glycerol, mannitol or erythritol, which act as compatible solutes in *A. nidulans* (Redkar *et al*., 1995; De Vries *et al*., 2003), to MM plates containing 0.5 M NaCl was unable to restore growth resembling that of the wild-type strain L20 and the control strains for the *stzA* deletion mutants or the *sltA1* strain GO281 (Figure 4.26).

Growth responses of all nine strains were also analysed on media imposing non-ionic osmotic stresses (Figure 4.27). All strains were similarly sensitive to MM containing 1.2 M sucrose and 1 M and 2 M sorbitol. All strains appeared resistant to 10% polyethylene glycol (PEG). The *stzA* deletion strains showed greater sensitivity to the presence of 6% ethanol (with 1% glucose as a carbon source) than the control strains and L20 (Figure 4.27). Ethanol acts as a chaotropic solute, disrupting the structure and function of macromolecules when compatible solutes are limiting (Hallsworth *et al*., 2003).

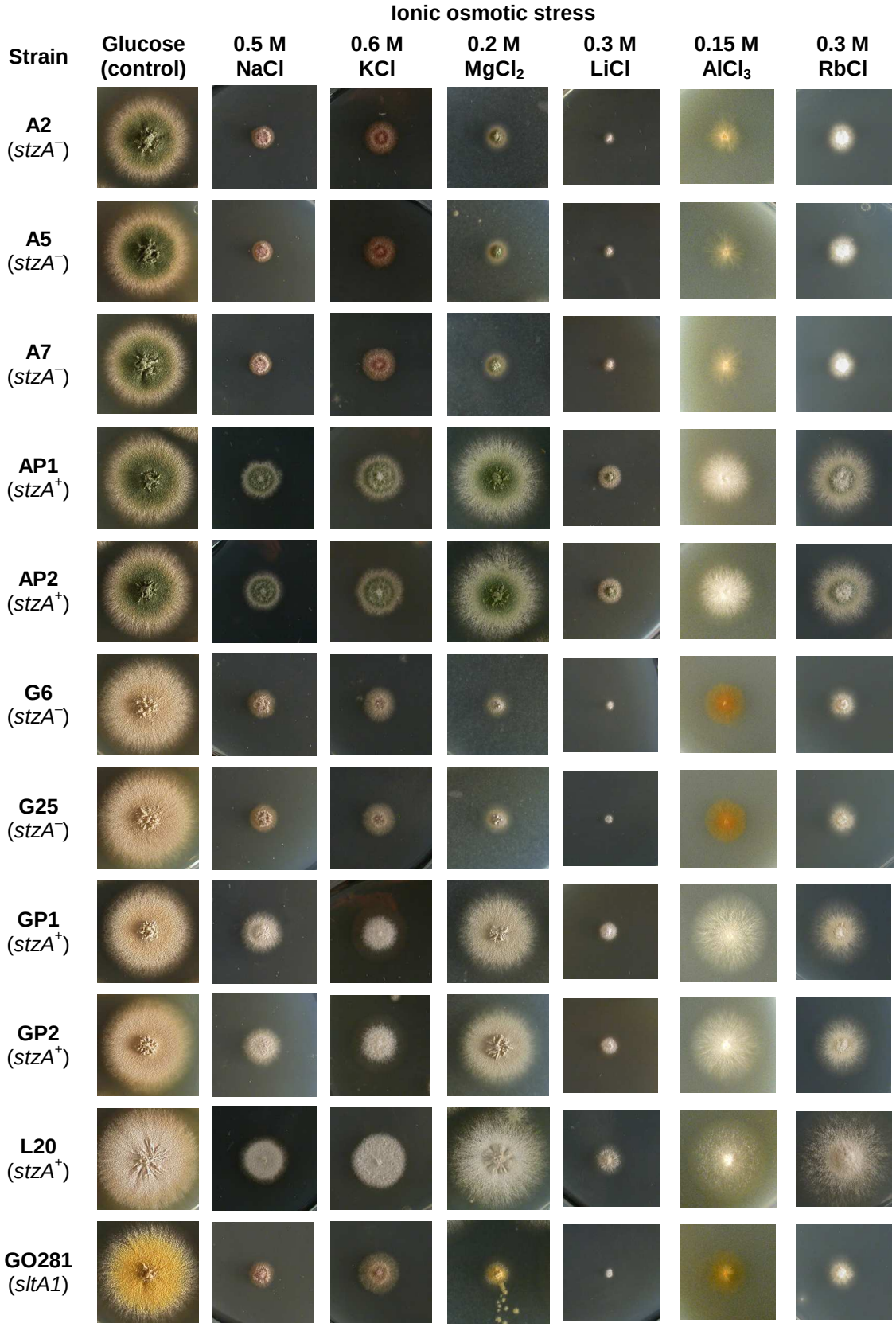


Figure 4.25. Growth responses of *stzA* deletants to ionic osmotic stresses. A photographic collage showing the growth responses of the *stzA* deletion strains (A2, A5, A7, G6 and G25); control strains transformed with the *pyr4* gene (AP1, AP2, GP1 and GP2); the wild-type strain L20; and the *sltA1* mutant GO281. Strains were grown on MM containing 1% glucose and 10 mM ammonium tartrate as the sole sources of carbon and nitrogen, respectively. All strains were incubated at 37 ºC for 48 hours.

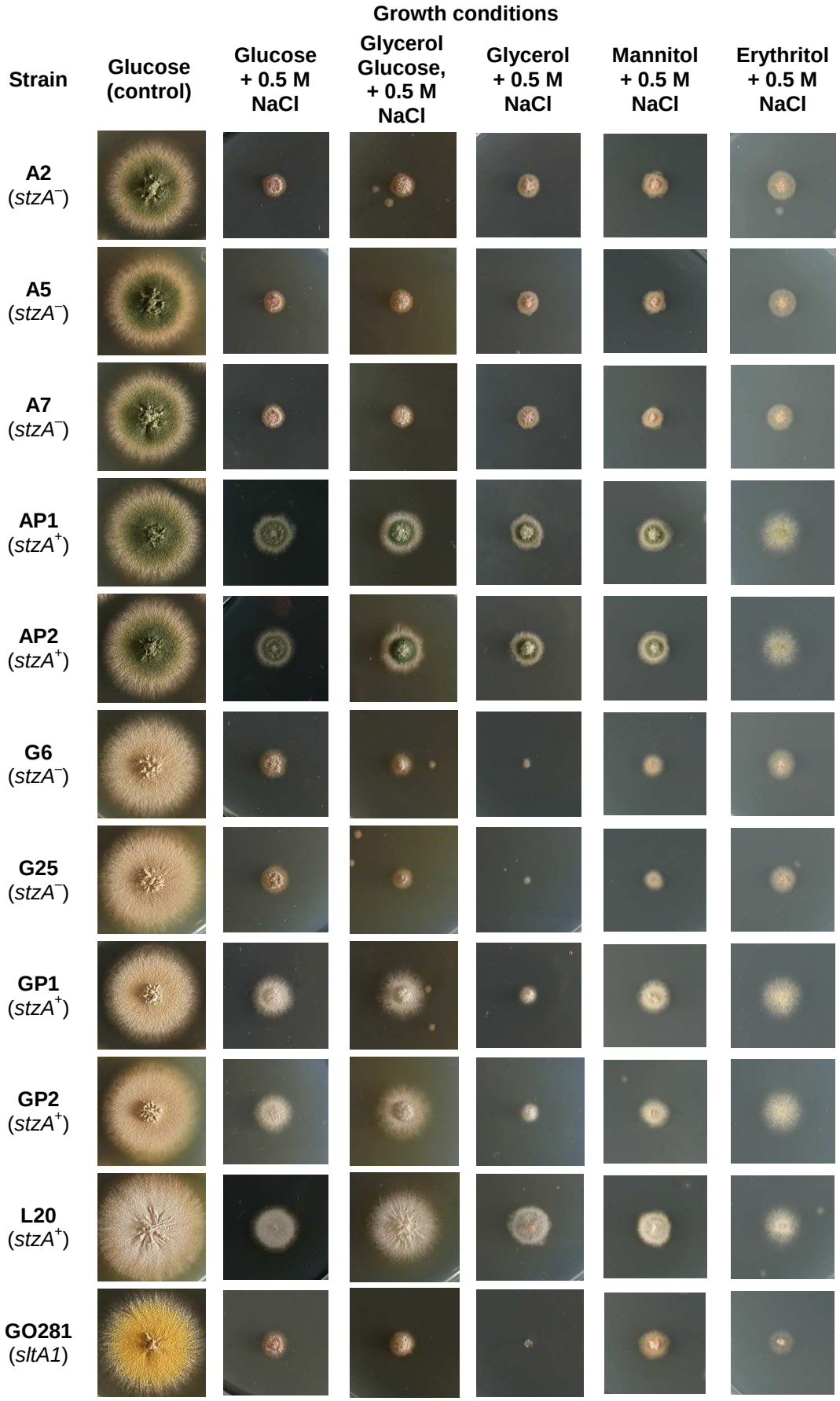


Figure 4.26. Growth responses of *stzA* deletants on MM containing 0.5 M NaCl supplemented with compatible solutes. All strains were incubated at 37 ºC for 48 hours. Poor growth of the control strains GP1 and GP2 on glycerol reflects the reduced ability of G191 to utilise glycerol as a sole carbon source (Section 4.3.3a)

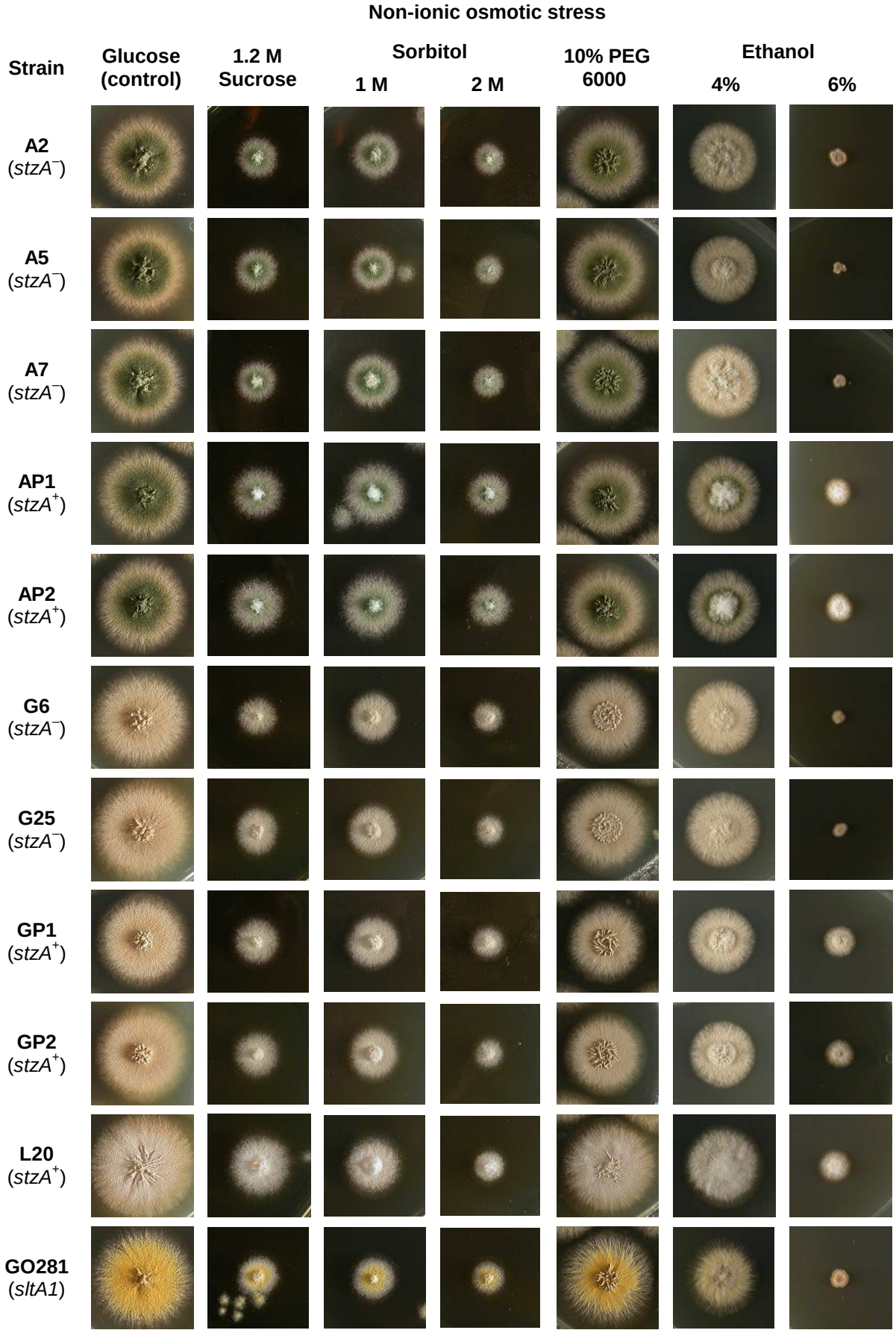


Figure 4.27. Growth responses of *stzA* deletants to non-ionic osmotic stresses. All strains were incubated at 37 ºC for 48 hours. For each treatment, glucose (1%) was used as a carbon source.

4.3.2 DNA damage responses

Responses of *A. nidulans stzA* deletants and *stzA*$^{+}$ strains to DNA-damaging agents on MEA plates were assessed by comparing growth morphologies in the presence of UV (6 mins exposure), the UV mimic 4NQO (4-nitroquiniline oxide) and the alkylating agent MNNG (*N*-Methyl-*N*'-Nitro-*N*-Nitrosoguanidine). The *stzA* deletants (A2, A5, A7, G6 and G25) showed a marked sensitivity to all DNA-damaging agents in a similar manner to GO281. In contrast the *sltA*$^{+}$ strains (AP1, AP2, GP1, GP2 and L20) were more tolerant to these DNA-damaging agents, shown by stronger radial growth under these conditions (Figure 4.28). Light treatment (2 hours) of *stzA* deletion strains post UV irradiation (6 mins) brought about a small but consistent recovery in growth for *stzA* deletants as well as *stzA*$^{+}$ strains.

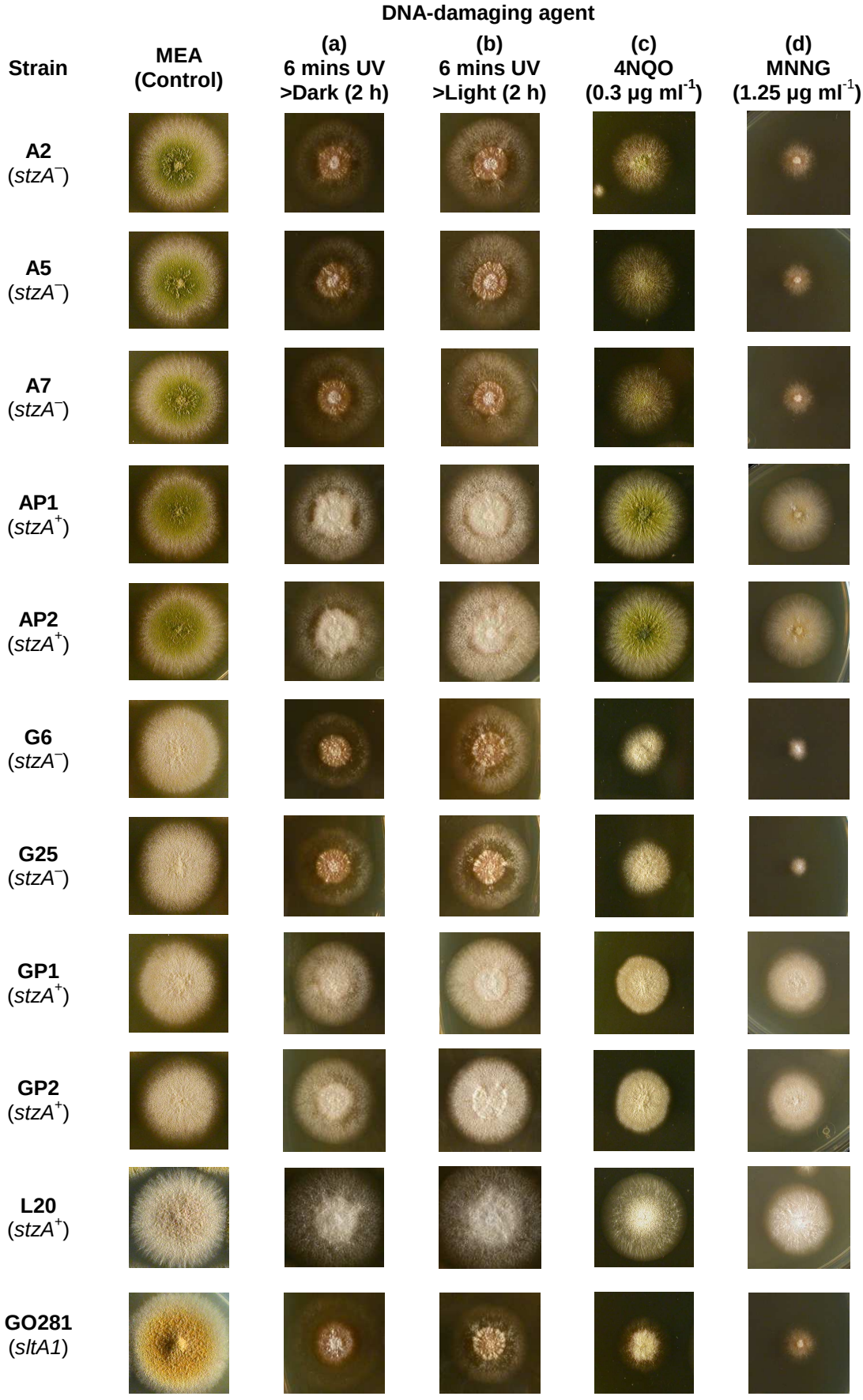


Figure 4.28. DNA-damage sensitivity tests. Strains were tested for their sensitivity to (a, b) UV radiation, (c) the UV mimic 4NQO and (d) the alkylating agent MNNG. Photoreactivation tests involved incubating colonies at 37 ºC for 24 hours before irradiating them with UV light for 6 minutes, and then left in conditions of (a) dark or (b) light for 2 hours before being returned to the incubator for a further 24 hours.

4.3.3 Carbon source utilisation

a) Glycolytic and gluconeogenic carbon substrate utilisation

The *stzA* deletants supported strong growth similar to the control strains and L20 on all the glycolytic carbon sources tested: glucose, fructose, sorbitol, mannitol, erythritol and galactose (Figure 4.29). Likewise, similar growth of all strains was supported by ethanol and the amino acids tested (glutamic acid, glutamine, proline, and leucine), which are gluconeogenic carbon sources (Figure 4.30). Glycerol supported strong growth of the A1149-derived transformants (the *stzA* deletants A2, A5 and A7 and the pDJB3-transformed controls AP1 and AP2). However, this carbon source supported weaker growth of G191-derived transformants (the *stzA* deletants G6 and G25 and the pDJB3-transformed controls GP1 and GP2). This is indicative of an uncharacterised mutation in G191 affecting the utilisation of glycerol.

All *stzA* deletants and GO281 grew poorly on the TCA cycle intermediates acetic acid, succinic acid, citric acid and also putrescine (Figures 4.30 and 4.31). These carbon sources supported better growth of the control strains and the wild-type strain L20. Growth of GO281 was even more retarded on these carbon sources because of its secondary mutation known to affect growth on gluconeogenic carbon sources (Chapter 3). Reduction of the growth of *stzA* deletants on these carbon sources could indicate that *stzA* plays a role in gluconeogenesis. However, it is perhaps more likely that these strains are sensitive to the considerable Na^+ concentrations of NaOH (equivalent to 0.13–0.16 M NaCl) required for pH adjustment of these media to pH 6.5. Increases in the relative growth of the *stzA* deletants compared to the control strains were observed when just 0.25% of acetic acid and succinic acid were used as sole carbon sources, which required substantially less NaOH to adjust the media to pH 6.5. However, when citric acid was used as a sole carbon source at a concentration of 0.25%, growth of *stzA* deletants remained poor compared to *stzA*$^+$ strains (Figure 4.31). Putrescine did not support strong growth of any strain when used as a sole carbon source at concentrations of 0.5% and 0.25% (results not shown).

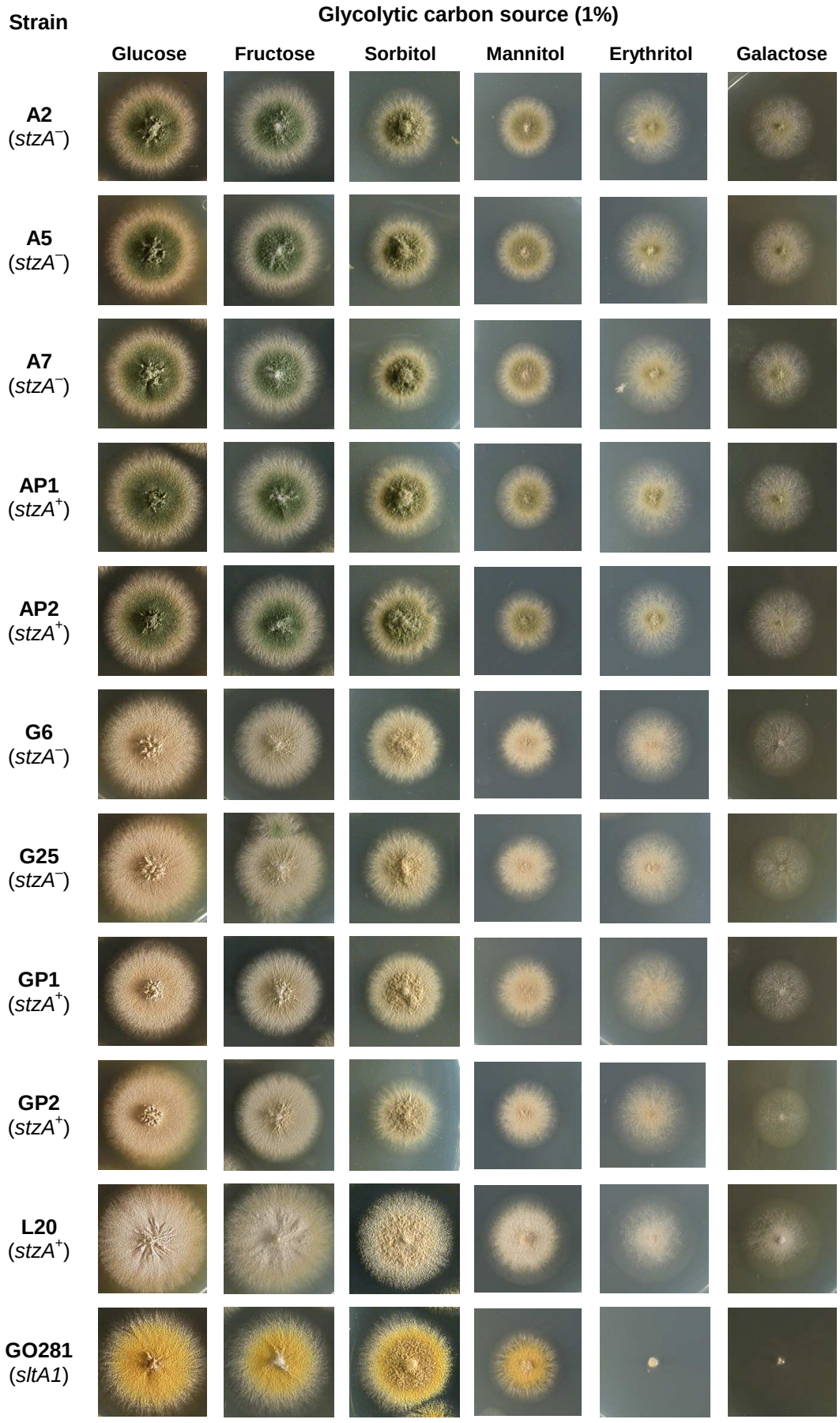


Figure 4.29. Utilisation of glycolytic carbon sources by *stzA* deletants. Ammonium tartrate (10 mM) was used as the nitrogen source. All strains were incubated at 37 ºC for 48 hours.

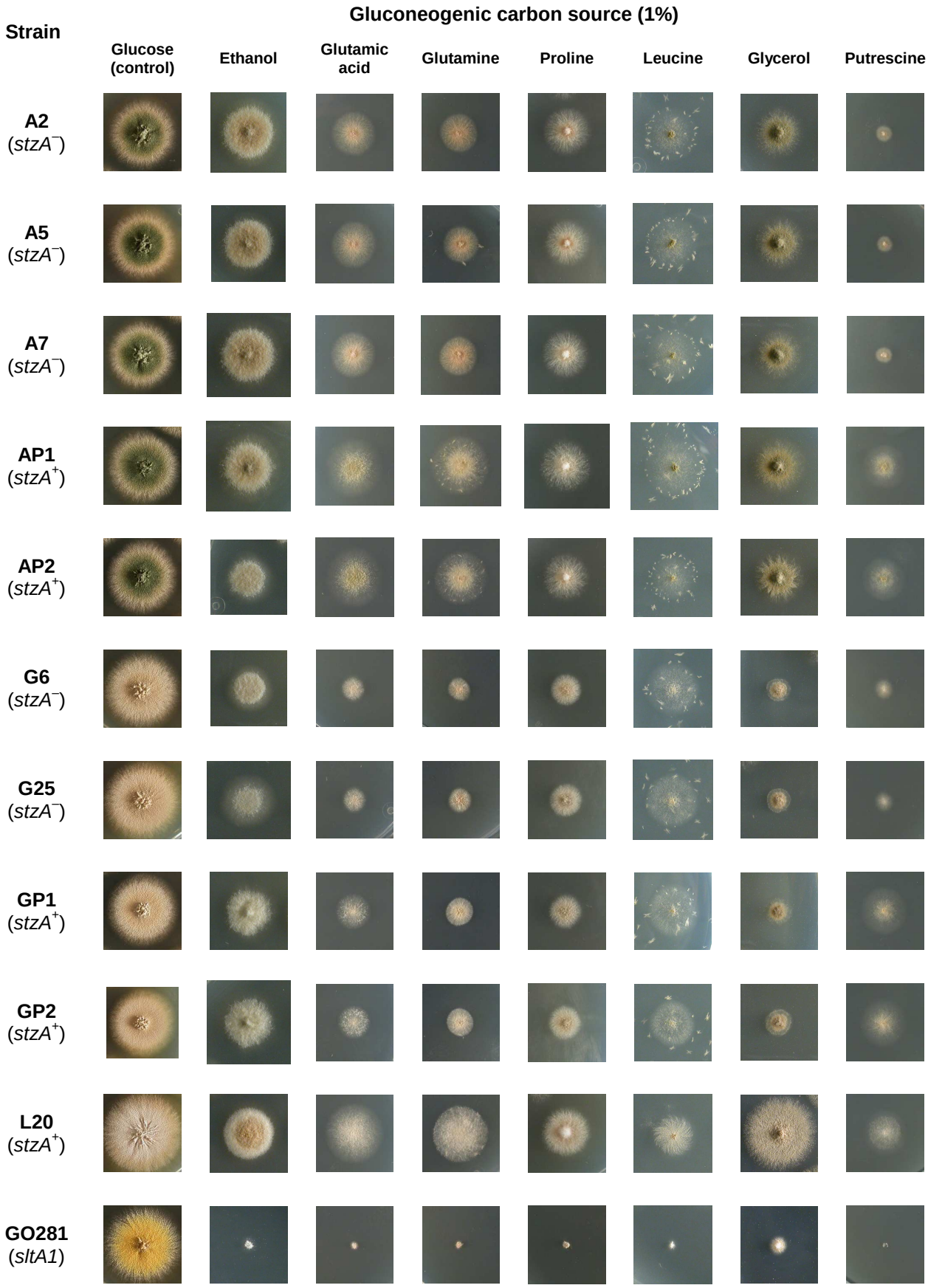


Figure 4.30. Utilisation of gluconeogenic carbon sources by *stzA* deletants. Ammonium tartrate (10 mM) was used as the nitrogen source. All strains were incubated at 37 ºC for 48 hours. Poor growth of the control strains GP1 and GP2 on glycerol is likely to reflect an uncharacterised mutation in G191 affecting utilisation of this carbon source.

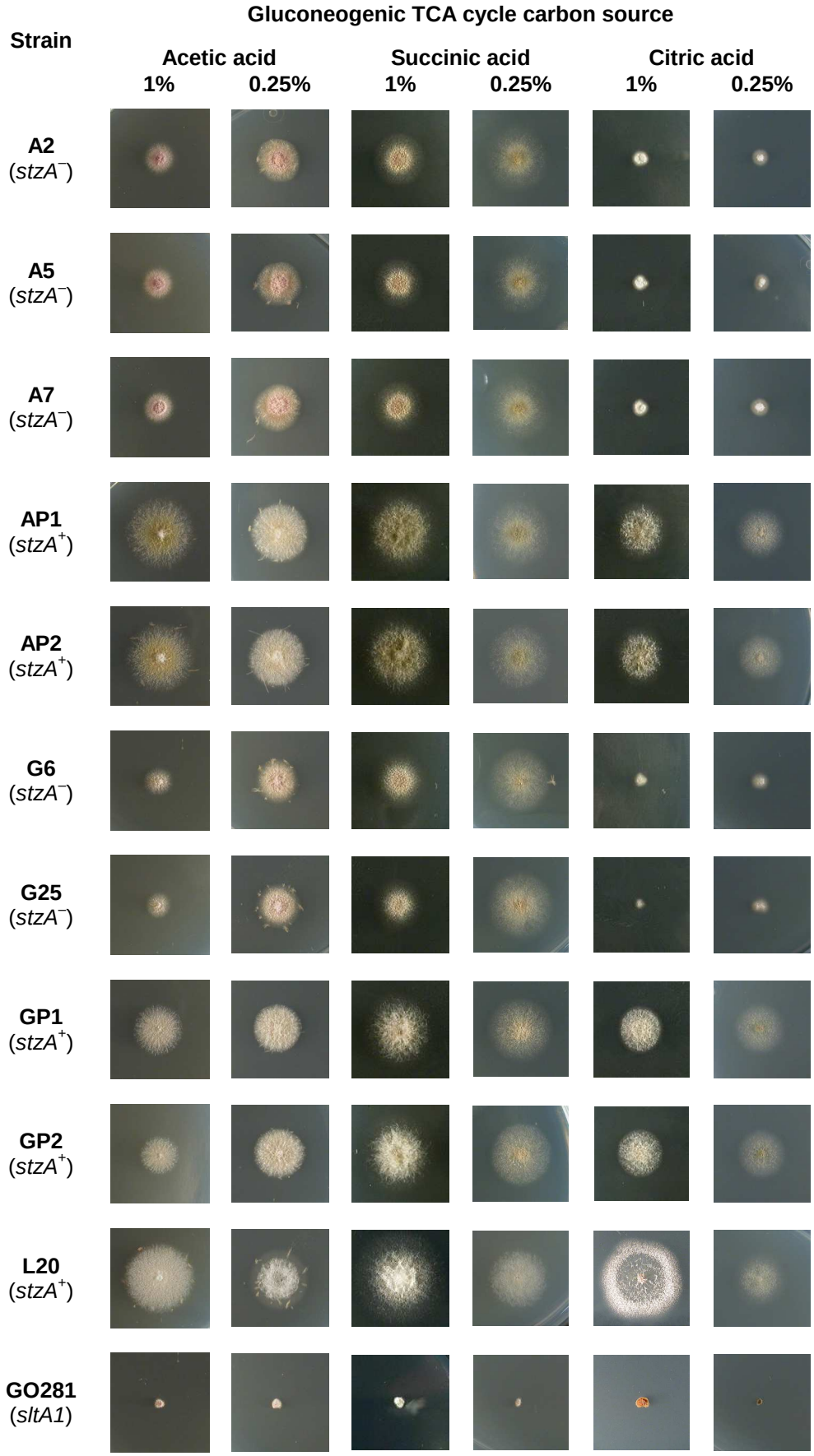


Figure 4.31. Utilisation TCA cycle intermediates as carbon sources by *stzA* deletants. Ammonium tartrate (10 mM) was used as the nitrogen source. All strains were incubated at 37 ºC for 48 hours.

b) Plant substrate utilisation

The growth of *stzA* deletants and control strains were compared when provided with plant substrates as sole carbon sources. Reduced growth of *stzA* deletion strains compared to *stzA*$^{+}$ strains was observed on pectin, polygalacturonic acid and galacturonic acid, three substrates that possess galacturonic acid as a backbone (Figure 4.32; De Vries *et al*., 2003). Note that poor growth of GP1 and GP2 on galacturonic acid is thought to be due to an uncharacterised mutation in G191 affecting its utilisation of this carbon source (and glycerol, shown in Figure 4.30). All strains grew well on starch and reasonably well on xylose and xylan. All strains grew rather poorly on cellulose as a carbon source after 2 days in the presence of either ammonium tartrate or urea, which are AreA-repressing and -de-repressing nitrogen sources, respectively (Lockington *et al*., 2002; results not shown).

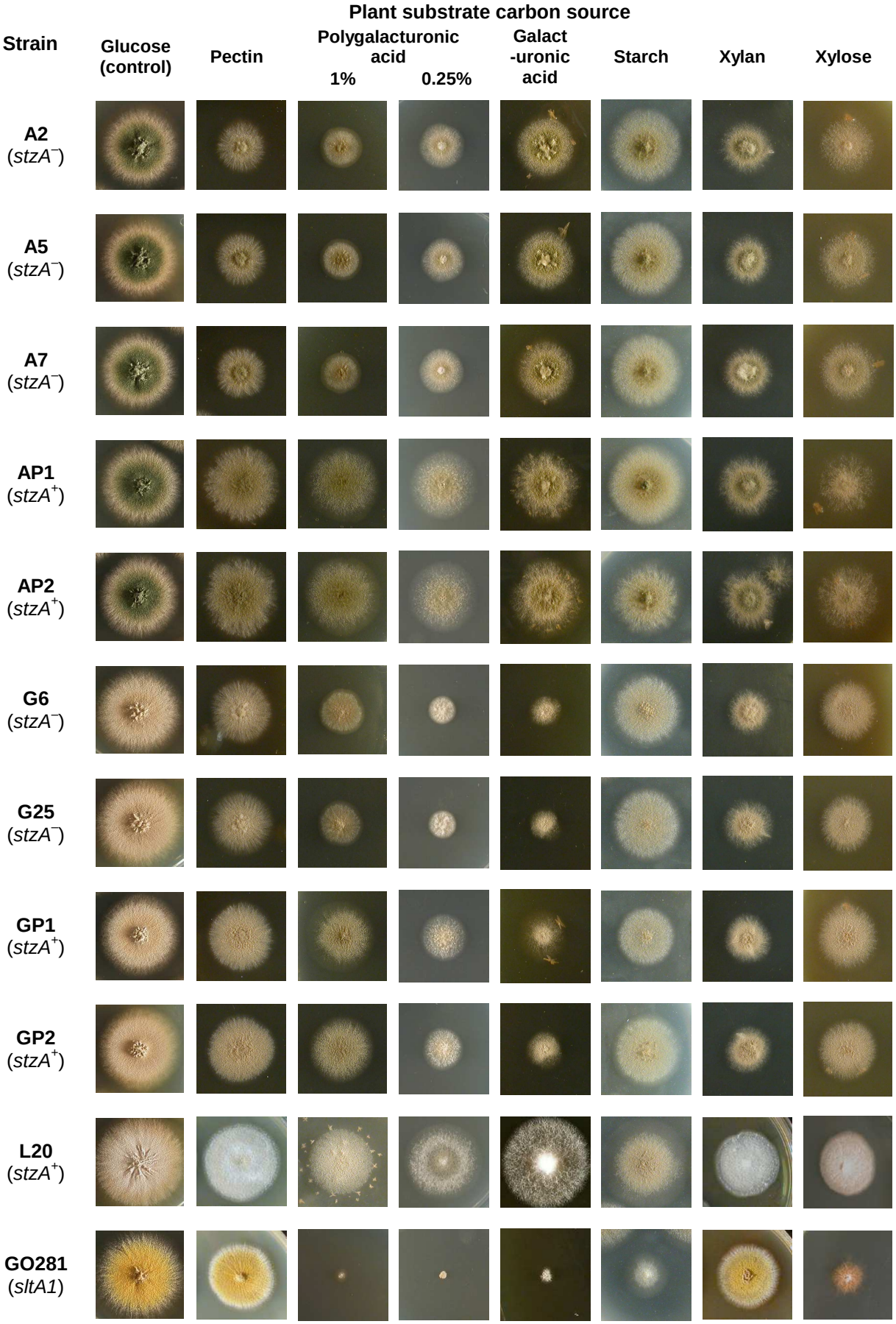


Figure 4.32. Utilisation of plant substrates as carbon sources by *stzA* deletants. All carbon sources were used at a concentration of 1% unless otherwise indicated. Ammonium tartrate (10 mM) was used as the nitrogen source. All strains were incubated at 37 ºC for 48 hours.

Considering the reduced growth of *stzA* deletants on pectin and polygalacturonic acid, it was decided to measure concentrations of polygalacturonase produced by the strains A2 (*stzA*$^{-}$) and A1149 (*stzA*$^{+}$) in liquid cultures (100 ml; 1 x 10^{7} spores ml^{-1}) with polygalacturonic acid as a sole carbon source after 48 hours of growth at 37 ºC. Enzyme concentrations were measured indirectly using the dinitrosalicylic acid (DNS) assay, which measures reducing sugars (in this case galacturonic acid levels). This method could also be used to indirectly measure cellulase and xylanase enzyme concentrations following growth of A2 and A1149 liquid cultures with cellulose and xylan as sole carbon sources. This was considered worthwhile considering that the *stzA* orthologue of *A. nidulans* in *T. reesei* (Ace1) is a known repressor of cellulase and xylanase expression (Aro *et al*., 2003). For the cellulase assay, urea was provided as the sole nitrogen source in the presence of cellulose as the sole carbon source. Ammonium was not used because AreA represses cellulase production in the presence of this nitrogen source, but such repression is relieved by urea (Lockington *et al*., 2002). Standard curves for glucose, xylose and galacturonic acid are shown in Figure 4.33. A standard curve for BSA protein using the Bradford assay is shown in Figure 4.34. It can be deduced that A2 and A1149 produced similar concentrations of cellulases, xylanases and polygalacturonases as similar amounts of reducing sugars were liberated per mg of protein for each sample (Tables 4.6–4.8).

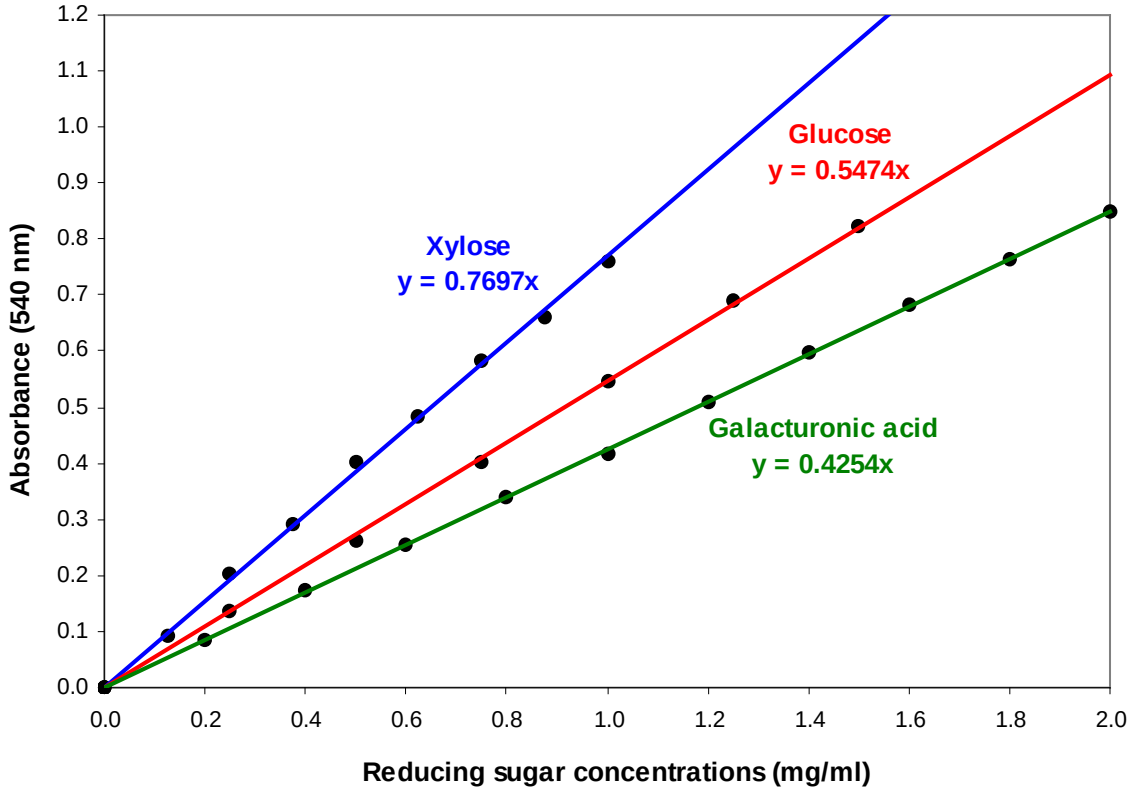


Figure 4.33. DNS assay standard curves for xylose, glucose and galacturonic acid.

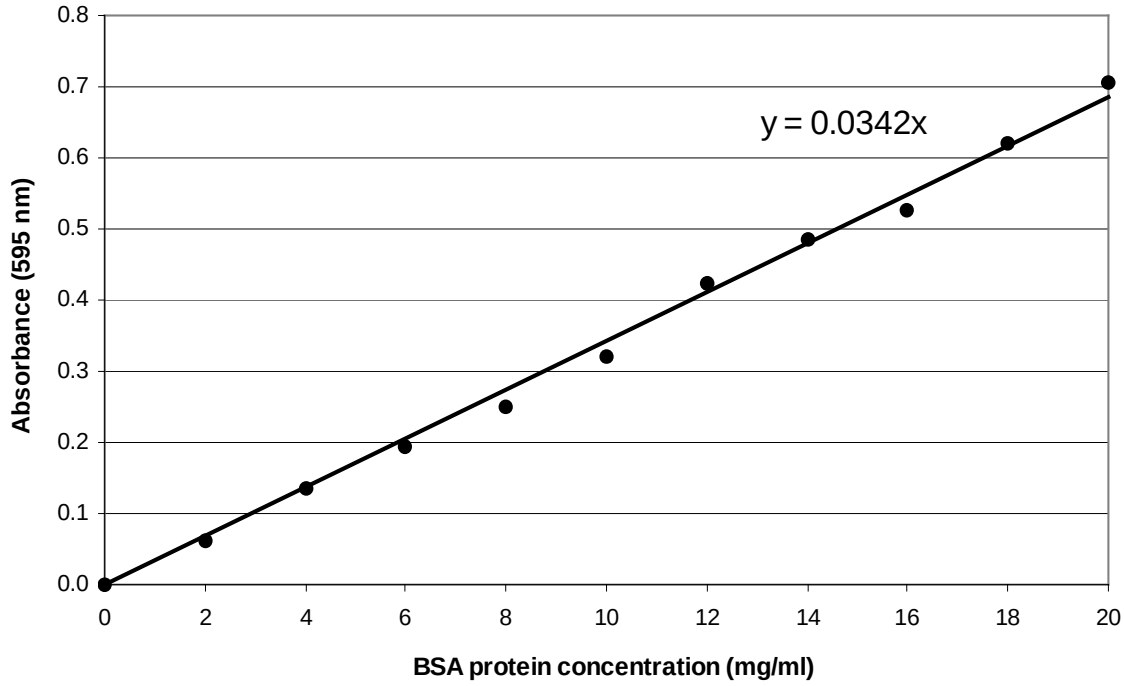


Figure 4.34. Bradford assay standard curve for BSA protein.

Table 4.6. DNS assay results on supernatants extracted from liquid cultures of *A. nidulans* strains A1149 and A2 (1 x 10^7 spores ml^{-1}) after 48 hours of growth at 37 ºC on 1% of each carbon source.

Carbon source	**Xylan**		**Cellulose**		**Polygalacturonic acid**	
Strain	**A1149**	**A2**	**A1149**	**A2**	**A1149**	**A2**
Mean A_{540} (4 replicates)	0.5733	0.5883	0.3320	0.3433	0.2380	0.2450
Dilution	10x	10x	10x	10x	10x	10x
Standard curve line equation	x = (y / 0.7697) x 10		x = (y / 0.5474) x 10		x = (y / 0.4254) x 10	
Reducing sugar liberated	Xylose		Glucose		Galacturonic acid	
Amount liberated per 100 µl sample in 30 mins (mg) (Means ± S.E., *n* = 4)	**7.45** ± 0.013	**7.64** ± 0.009	**6.07** ± 0.016	**6.27** ± 0.020	**5.59** ± 0.017	**5.76** ± 0.022

Table 4.7. Total protein concentrations determined using the Bradford assay.

Carbon source	**Xylan**		**Cellulose**		**Polygalacturonic acid**	
Strain	**A1149**	**A2**	**A1149**	**A2**	**A1149**	**A2**
Mean A_{595} (4 replicates)	0.3858	0.3910	0.2155	0.2215	0.5348	0.5280
Dilution	4x	4x	Undiluted	Undiluted	2x	2x
Standard curve line equation	x = y / 0.0342		x = y / 0.0342		x = y / 0.0342	
Protein concentration (mg / ml)	45.20	45.85	6.27	6.52	31.29	30.85
Protein concentration per 100 µl sample (mg) (Means ± S.E., *n* = 4)	**4.51** ± 0.008	**4.57** ± 0.010	**0.630** ± 0.003	**0.648** ± 0.003	**3.13** ± 0.004	**3.09** ± 0.005

Table 4.8. Amounts of protein liberated per mg of protein for each 100 µl sample.

Carbon source	**Xylan**		**Cellulose**		**Polygalacturonic acid**	
Reducing sugar liberated / mg of protein for 100 µl sample (mg) (Means ± S.E.)	**1.65** ± 0.013	**1.67** ± 0.014	**9.63** ± 0.163	**9.68** ± 0.159	**1.79** ± 0.017	**1.86** ± 0.022

4.3.4 Nitrogen source utilisation

All nine strains grew equally well on ammonium tartrate, sodium nitrate, glutamine, glutamic acid, proline and phenylalanine when used as sole nitrogen sources (10 mM; Figure 4.35). The *stzA* deletion mutants were also tested for their sensitivity to arginine. All *stzA* deletion mutants were sensitive to arginine (50 mM) when present as a sole nitrogen source (pH 8.0), and showed a similar phenotype to the *sltA1* mutant GO281, notable for its characteristic red pigment production and reduced radial growth under these conditions (Clement *et al*., 1996; O'Neil *et al*., 2002). Although this stressed phenotype was observed at pH 8.0, it was not observed at pH 6.5.

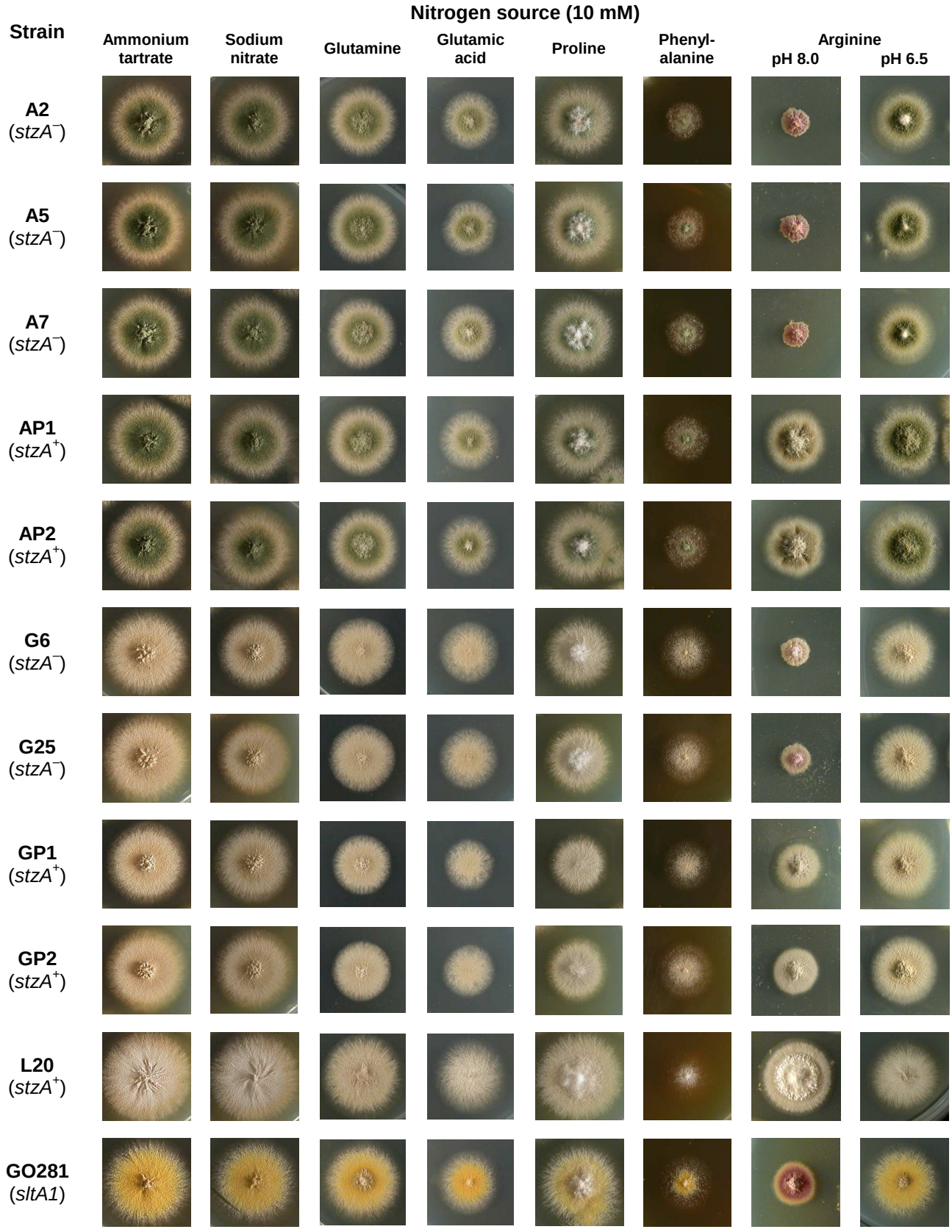


Figure 4.35. Nitrogen source utilisation by *stzA* deletants. Nitrogen sources were used at a concentration of 10 mM, with the exception of arginine (50 mM). Glucose (1%) was used as the sole carbon source. All strains were incubated at 37 ºC for 48 hours.

4.3.5 pH responses

The *stzA* deletants demonstrated similar growth responses at pH 5.0, 6.5 and 8.0 when compared to *stzA*$^+$ strains (Figure 4.36). However, when the MM at pH 8.0 was buffered with 100 mM Na_2HPO_4, substantially reduced growth of the *stzA* deletants and GO281 was observed. This phenomenon was observed by Spielvogel *et al*. (2008), but growth of strains in unbuffered medium at pH 8.0 was not tested.

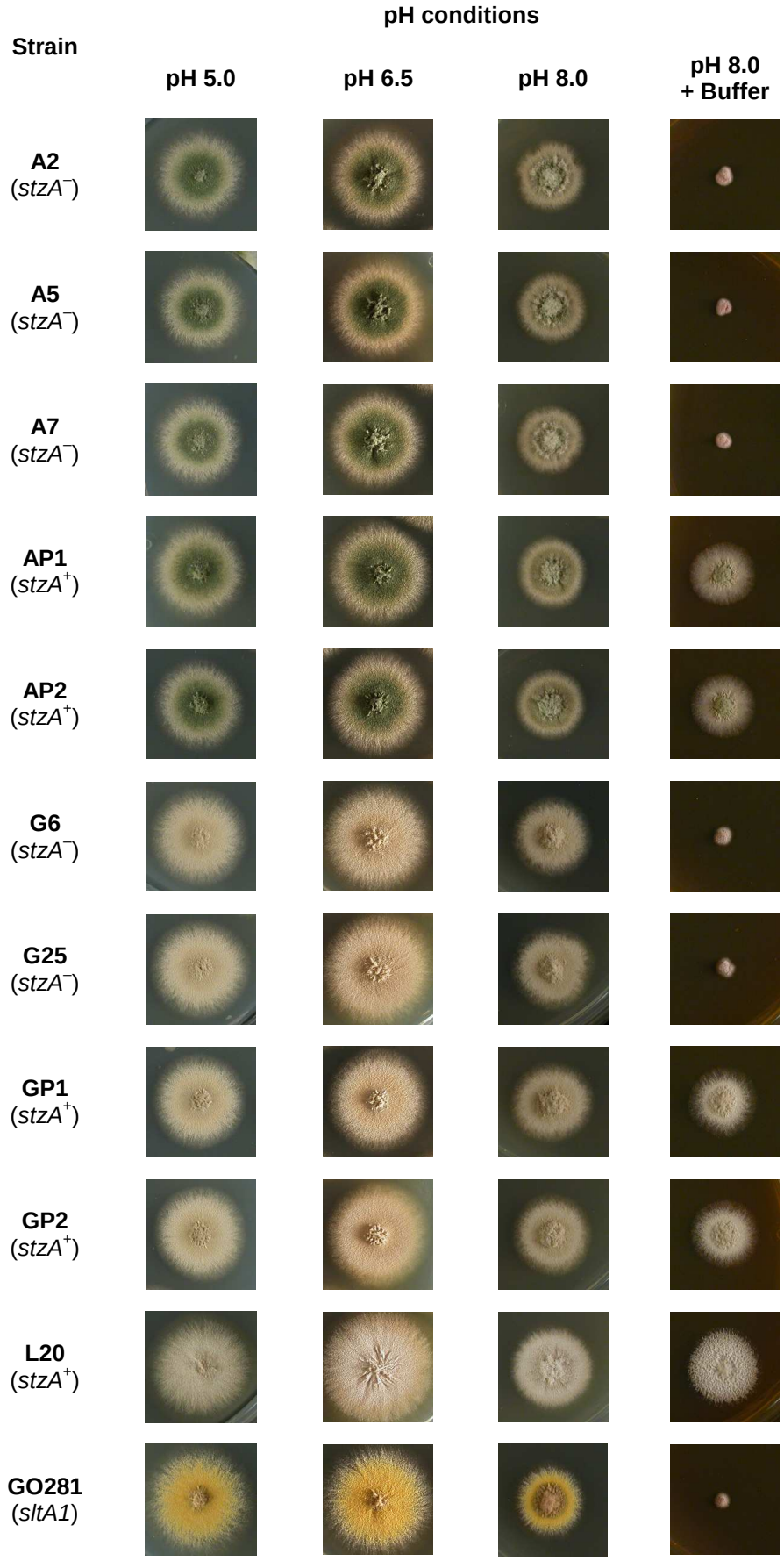


Figure 4.36. Growth responses of *stzA* deletants to different pH. All strains were incubated at 37 ºC for 48 hours. Glucose was used as the sole carbon source. The buffered MM at pH 8.0 was buffered with 100 mM Na_2HPO_4 used by Spielvogel *et al*. (2008).

4.3.6 Oxidative stress and fungicide responses

The *stzA* deletants and *stzA*$^{+}$ strains were equally sensitive to the oxidative stresses exerted by hydrogen peroxide and menadione (Figure 4.37) and also to the fungicides congo red, calcofluor white, amphotericin B and fludioxonil (Figure 4.38). However, *stzA* deletion strains are known to exhibit much greater sensitivity to neomycin than wild-type strains (Spielvogel *et al*., 2008), results that are reflected in the present study.

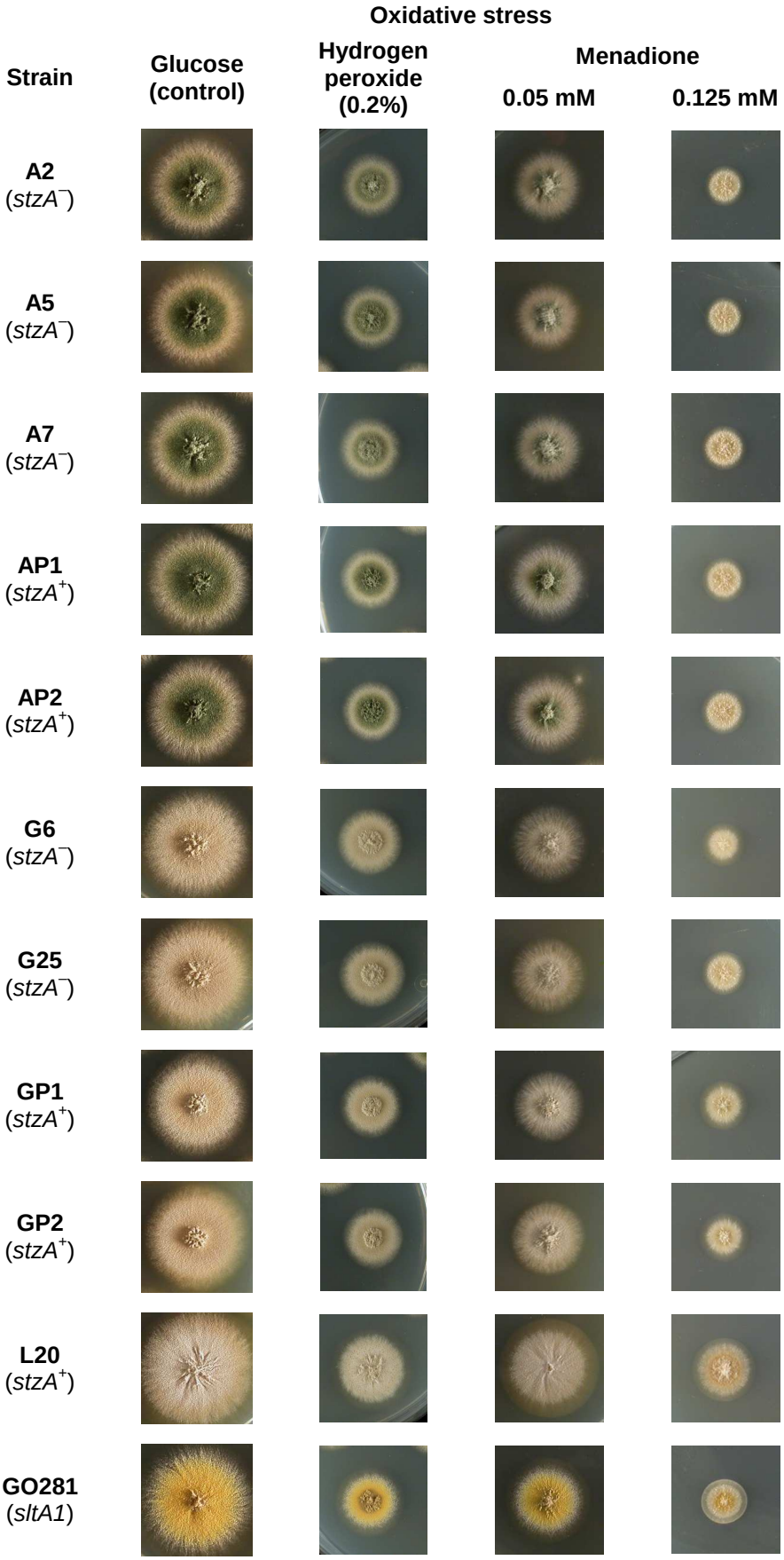


Figure 4.37. Growth responses of *stzA* deletants to oxidative stresses. All strains were incubated at 37 ºC for 48 hours. Glucose was used as the sole carbon source.

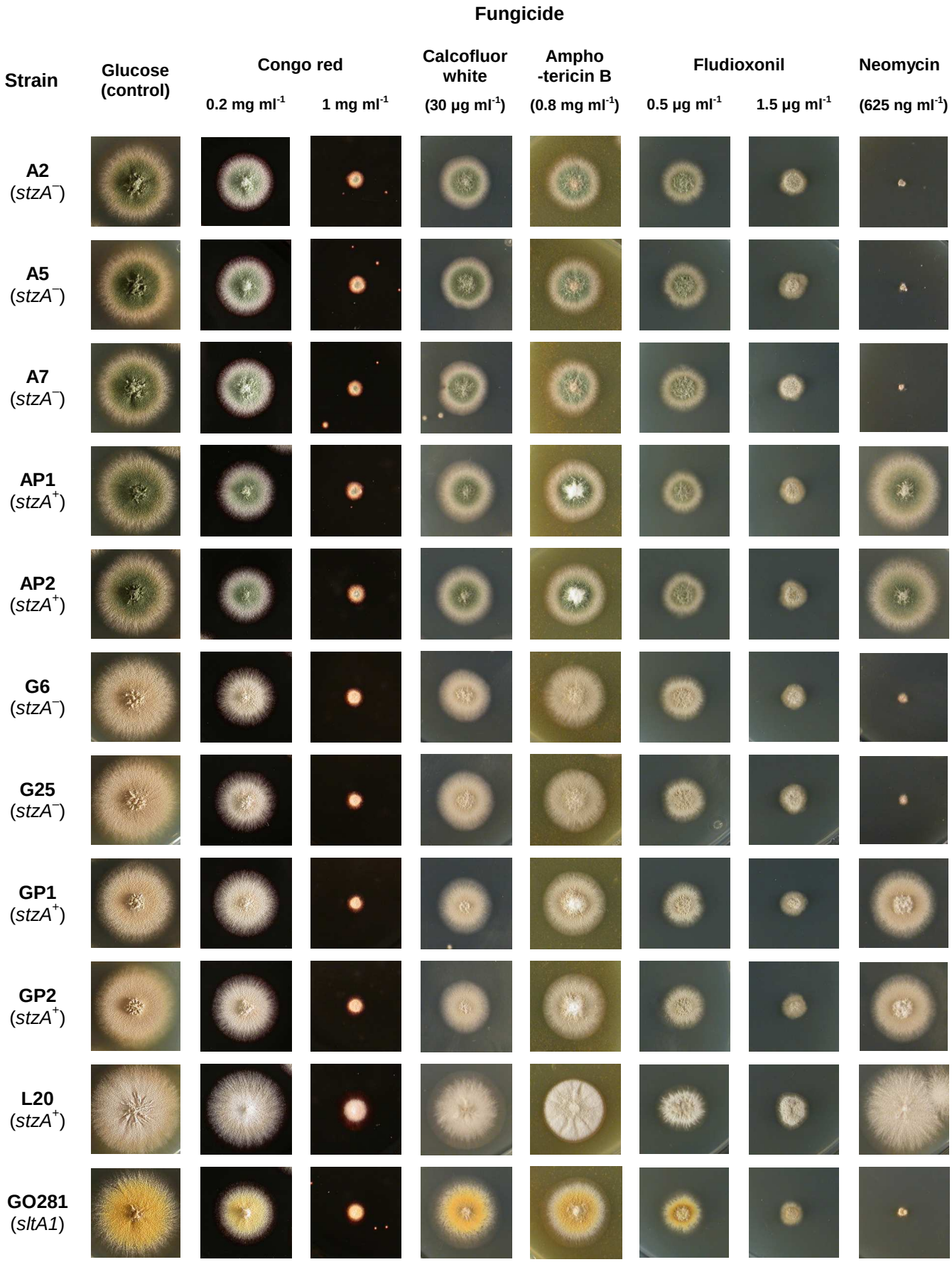


Figure 4.38. Growth responses of *stzA* deletants to fungicides. All strains were incubated at 37 ºC for 48 hours. Glucose was used as the sole carbon source.

4.3.7 *T. reesei ace1* and wild-type stress responses

The *T. reesei* wild-type (QM9414) and *ace1* deletion (VTT-D-061244) strains exhibited similar growth responses to the osmotic stresses NaCl and KCl; the DNA-damaging agents UV, 4NQO and MNNG; 50 mM arginine as a sole nitrogen source; neomycin; and acidic, neutral and alkaline pH (Figure 4.39). The *ace1* deletion strain is known to exhibit impaired growth on sorbitol (2%) as a sole carbon source (Aro *et al*., 2003), and therefore this phenotype could be used as a control.

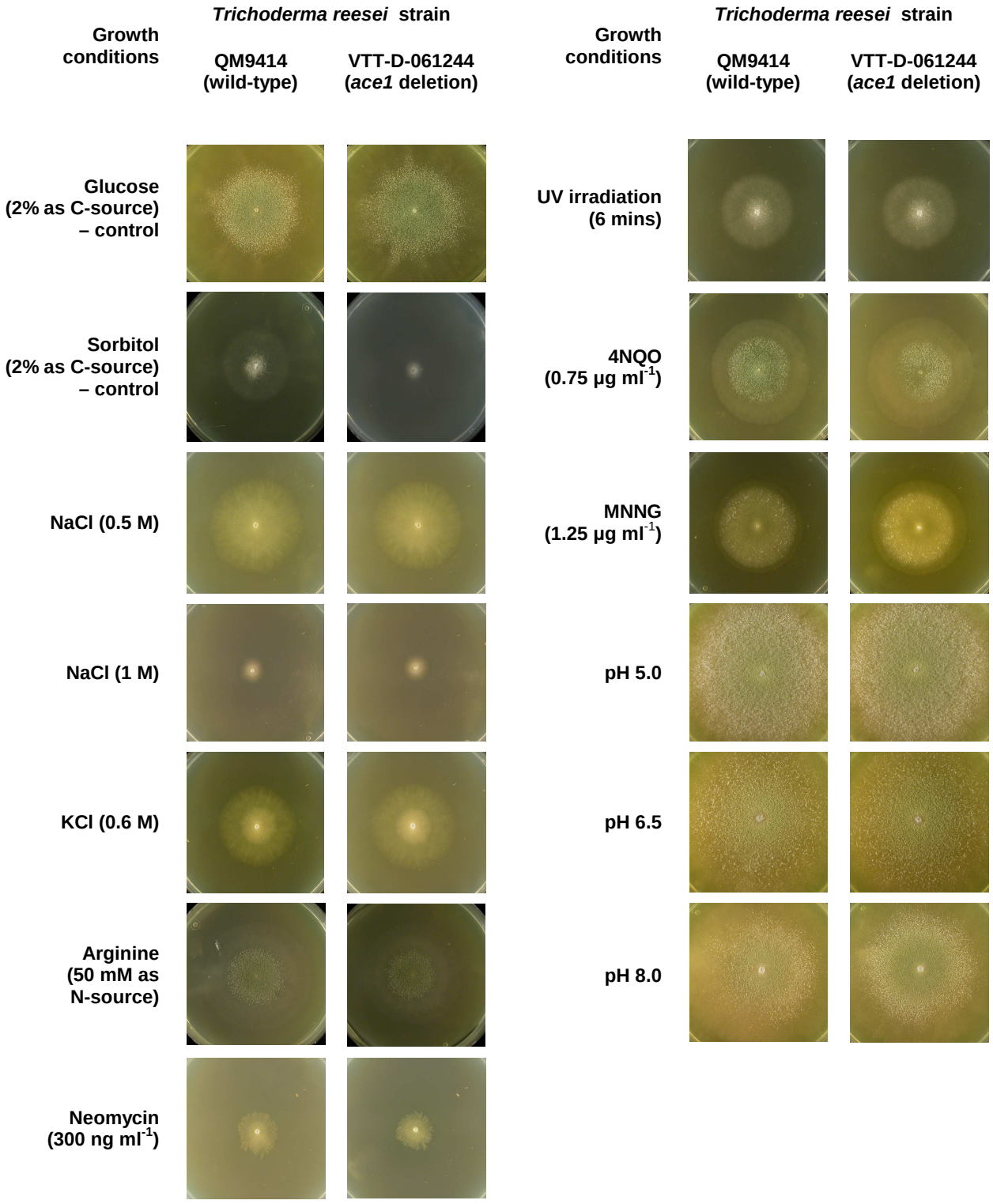


Figure 4.39. Growth responses of the *T. reesei* wild-type and *ace1* deletion strains to various conditions of nutrition and stress. Strains were grown on MEA (pH 6.5) or with glucose (2%) and ammonium tartrate (10 mM) as sources of carbon and nitrogen, respectively, unless otherwise indicated. Strains were incubated at 30 ºC for 72 hours.

4.4 Discussion

4.4.1 Amplification of *stzA* gene-targeting substrates

The development of fusion PCR techniques and the availability of *ku* deletion strains provide a powerful system for gene deletion in fungi (reviewed in Kück and Hoff, 2010), which can be realised with the availability of complete genome sequences for fungi that include *A. nidulans* (Galagan *et al*., 2005; Wortman *et al*., 2009). Deletion of the *stzA* ORF in *A. nidulans* was attempted using the gene replacement method devised by Nielsen *et al*. (2006). This method involved the generation of two bipartite gene-targeting substrates by fusing DNA sequences that flank the *stzA* target gene with *pyr4* gene fragments. The substrates were then used to transform *A. nidulans nkuA*$^{-}$ (A1149) and *nkuA*$^{+}$ (G191) strains (Figure 4.2). Oligonucleotide sequences used for the project are shown in Table 4.1.

The initial gene fragments were amplified by normal PCR. Portions of the *pyr4* gene (designated P1 and P2) were easily amplified by PCR using pDEL1 as a template (Figures 4.3 and 4.7). Likewise, the *stzA*-flanking regions (designated U and D) were easily PCR-amplified from *A. nidulans* L20 genomic DNA (Figures 4.5, 4.6 and 4.8a). These four fragments were synthesised using the PCR conditions shown in Table 4.2. The ease with which *stzA*-flanking regions were amplified may be attributed to the decision to synthesise short *stzA*-flanking regions (970 and 802 bp). Visible non-specific amplification products, such as those amplified along with the P1 fragment, were successfully removed by DNA band purification (Figure 4.8b).

Fusion PCR was used to generate the two bipartite *stzA* gene-targeting substrates. Attempts to synthesise fusion product 1 (2,508 bp) from fragments P1 and U using the primers U1 and M3 proved extremely difficult. Optimisation of key fusion PCR parameters, such as primer annealing temperature, $MgCl_2$ concentration, DNA concentration and fusion cycle number, failed to produce any discernible fusion product (Figures 4.9 and 4.10). Furthermore, optimisation of these parameters resulted in only low yields of fusion product 2 (2,321 bp) from the fragments P2 and D when using the primers M2 and D2 (Figures 4.12 and 4.13).

Amplification using nested primers, as opposed to the original primers, often greatly increases yields of the desired fusion product by increasing the specificity of amplification and therefore also reduces the possibility of obtaining non-specific amplification products (Szewczyk *et al*., 2006). The use of nested fusion PCR was

key in producing high yields of fusion products 1 and 2 using primer pairs F1 and F2, and F3 and F4. Because the new primers were further from the 3' and 5' termini of the fused PCR products, these products were shortened to 2,321 and 1,944 bp (Table 4.3). High yields of fusion product 1 could be generated by using a primer annealing temperature of 58 ºC, an $MgCl_2$ concentration of 2.5 mM, 10 fusion cycles and DNA fragments present in concentrations of 100 ng / 50 µl (Figure 4.11). High yields of fusion product 2 could be generated by using a primer annealing temperature of 54 ºC, an $MgCl_2$ concentration of 1.5 mM, 5 fusion cycles and DNA fragments present in concentrations of 100 ng / 50 µl (Figure 4.14). PCR extension times were selected to allow 1 minute per 1 kb of elongation and at least an additional 2 minutes. Furthermore, a final extension time of 10–15 minutes allowed for the complete extension of all amplification products (Sambrook and Russell, 2001). Although annealing temperatures were maintained at 55 ºC for both sets of fusion tags (as advised by Uffe Mortensen, personal communication), annealing temperatures for the primer pairs were optimised independently for fusion products 1 and 2 and were adjusted to 58 ºC and 54 ºC, respectively. The range of temperatures used for primer annealing optimisation was based on the "Wallace rule" used to estimate the melting temperatures of primers and their complementary target sequences: T_m (ºC) = 2 (A+T) + 4 (G+C) (Sambrook and Russell, 2001). The optimised fusion PCR conditions are summarised in Table 4.4.

The effect of nested PCR was to shorten *stzA*-flanking fragments (from 970 to 849 bp; and 802 to 721 bp) used for HR with *A. nidulans* genomic DNA during transformation (Figures 4.9c and 4.12c). These regions were much shorter than the flanking regions (2033 and 2088 bp) used to delete the *radC* gene by Nielsen *et al*. (2006). It was decided to use shorter flanking regions because shorter DNA fragments are generally easier to amplify by PCR and the chance of a mutation being created during PCR reactions is reduced (Sambrook and Russell, 2001). Furthermore, it has been demonstrated that flanking regions of just 500 bp are sufficiently long to allow efficient gene targeting in an *nku70* deletion strain of *A. nidulans* (Nayak *et al*., 2006). Gene targeting was similarly efficient (between 89–92%) with 2 kb, 1 kb and 500 bp flanking regions although fewer transformants were obtained (60, 50 and 36, respectively) upon deletion of the histone H1 gene. In *N. crassa*, homology arms of 500 bp resulted in only a 10% reduction in correct transformants compared to the use of 1 kb fragments (Ninomiya *et al*., 2004). Use of the nested PCR approach also shortens the *pyr4* gene fragments, but no gene sequence is lost because the two shorter *pyr4* sequences still overlap, and HR can still occur between these sequences.

All PCRs were performed with Phusion® DNA polymerase (Finnzymes). Use of an enzyme that generates blunt ends in the amplification products was essential so that the correct reading frame of the *pyr4* selectable marker would be maintained following HR. Furthermore, Phusion® DNA polymerase has higher fidelity compared to *Taq* DNA polymerase because it possess 3'–5' exonuclease proofreading activity in addition to 5'–3' DNA polymerase activity. Proofreading activity helped to prevent mutations in the amplified *pyr4* gene fragments. The final concentration of Phusion® DNA polymerase (0.01 U μl^{-1}) was not increased during the PCR optimisation process. Maintaining a lower concentration of the enzyme in PCR reactions reduces the incidence of high molecular weight DNA smears. Due to the increased processivity of Phusion® DNA polymerase, there was no advantage in increasing dNTP concentrations, so they were maintained at 0.2 mM for optimum results (Phusion® DNA polymerase handbook).

Four separate single-enzyme restriction digests were performed for both fusion products 1 and 2. All restriction fragments were of the expected sizes, indicative of the desired sequences having been amplified and fused successfully (Figure 4.16). Additional fainter bands present on the gels were likely due to restriction of non-specific PCR products present in high molecular weight DNA smears. These non-specific products were removed prior to transformation by running the PCR samples on a gel and by extracting and purifying the DNA bands containing the desired fusion products (Figure 4.15).

4.4.2 Deletion of the *stzA* gene

Transformation in *A. nidulans* was optimised using 2 x 10^7 protoplasts of the *pyrG*89 (mutant *pyrG*) strains G191 and A1149 using 1 µg of the plasmid pDJB3 (Figure 4.4; Appendix 4). This plasmid contains the *pyr4* gene (encoding orotidine-5'-phosphate decarboxylase) from *N. crassa* and can complement the *pyrG89* mutation of its orthologue in *A. nidulans*. The presence of the 3.5 kb *ans1* sequence within the vector, which is reiterated in the *A. nidulans* genome, is known to dramatically increase the transformation efficiency of this vector due to HR events. Accordingly, plasmids containing *ans1* can transform G191 at a frequency of ~5000 stable transformants per µg of DNA (Ballance and Turner, 1985). Complementation can also occur by ectopic integration into the *A. nidulans* genome as long as the integration event does not disrupt a gene essential for viability on the selective regeneration medium (Oakley *et al*., 1987).

In the present study, G191 was used to optimise transformation efficiencies with the plasmid pDJB3 (Figure 4.17). Transformation efficiencies were increased from an average of 14 transformants to 36 per µg of plasmid DNA by reducing the concentration of agar from 1.5% to 0.5% in the regeneration overlay medium. The survival of protoplasts was likely enhanced by being able to plate protoplasts into the overlay medium at a reduced temperature before the agar solidified, owing to the reduced agar concentration. Sudden exposure to temperatures of approximately 46 ºC is known to cause a severe heat shock to *A. nidulans* protoplasts and approaches the lethal temperature for their survival (Newbury and Peberdy, 1996). The transformation efficiency could be increased further to an average of 57 transformants per µg of plasmid DNA by plating the transformants directly onto the regeneration medium rather than into an overlay medium, thus avoiding heat shock completely. By combining this method of direct plating with the addition of 0.6 M KCl to the PEG solution, approximately 250 transformants could be obtained for both G191 and A1149. It should be noted that the PEG solution used by Tilburn *et al*. (1983) lacked 0.6 M KCl. The molarity of this solution is approximately 0.1 M and the addition of 0.6 M KCl has been shown to increase transformation efficiencies considerably by reducing osmotic shock-induced protoplast lysis (Oakley *et al*., 1987). As similar numbers of transformants were obtained for G191 and A1149 using pDJB3, it is likely that the vast majority of pDJB3 integrations were homologous in both strains because ectopic integration is substantially minimised in Ku-deficient strains such as A1149.

Two stable transformants obtained with G191 and A1149 using pDJB3 (designated GP1, GP2, AP1 and AP2) were used as controls for each strain in subsequent phenotypic analyses. These control strains were randomly selected from the transformants obtained with pDJB3 and were considered suitable controls because these, like the *stzA* deletion transformants, had been through the transformation process and acquired *pyr4* to complement the *pyrG89* mutation, but still possessed a functional *stzA* gene. Hence, the controls and the *stzA* deletion transformants were essentially isogenic, originating from either A1149 or G191, and differed only in *stzA* ORF presence/absence and the locus of *pyr4* integration.

Using the optimised transformation method, the two bipartite gene-targeting substrates – fusion products 1 and 2 (200 ng) – were used subsequently to target the *stzA* gene *in vivo* by HR using transformation with 2 x 10^7 protoplasts of A1149 and G191 strains. Eleven and 28 transformants (labelled A1–A11 and G1–G25) formed colonies on the regeneration medium for A1149 and G191, respectively. However, only 3/11 and 2/28 of these transformants (designated A2, A5, A7, G6 and G25) regenerated when re-plated onto selective

medium. Hence, 72.7% and 92.9% of transformants for A1149 and G191 were abortive. The absence of the *nkuA* gene in strains such as A1149, which suppresses NHEJ and facilitates HR of DNA fragments into the *A. nidulans* genome (Krappmann, 2007), may explain why fewer transformants in this strain were abortive than those of G191. Hence, in A1149, it seems more likely that the integration of the homologous gene-targeting substrates at the *stzA* locus was facilitated and the intregration of these fragments at other regions was suppressed, preventing the disruption of essential genes.

The selectable *pyr4* marker used for transformation with either plasmid or *stzA* gene-targeting substrates, came from the pDJB3 and pDEL1 plasmids, respectively, and in both cases originated from *N. crassa*. Use of a selectable marker that was not an *A. nidulans* gene was important. If *A. nidulans pyrG* is used, the copy of the transforming DNA fragment can repair the mutant chromosomal copy through a double crossover HR event and complement uridine/uracil deficiency without the occurrence of *stzA* deletion (Szewczyk *et al*., 2006). Crucially, no evidence could be found for the stable integration of *N. crassa pyr4* at the *A. nidulans pyrG* locus (Ballance and Turner, 1985)

A PCR assay was designed to demonstrate that (a) one of the two fusion products (fusion product 2) used as an *stzA* gene-targeting substrate was derived from its correct substrate fragments and that (b) the fusion product had integrated at the *stzA* locus during *A. nidulans* transformation (Figure 4.19 and Table 4.5). The expected size of the PCR product was 2,437 bp and could be generated by the primer pair S1 and S2. The PCR assay product was obtained from the genomic DNA of all five potential *stzA* deletion transformants (A2, A5, A7, G6 and G25) but not that of untransformed A1149 used as a control (Figures 4.21 and 4.22a). Restriction of the PCR assay product obtained from A2 yielded fragments of the expected sizes when restricted with *Bam*HI alone and *Bam*HI and *Xba*I in a double digest (Figure 4.22b and c).

Sequence data was obtained by the Dundee Sequencing Centre (http://www.dnaseq.co.uk) for the A2 PCR assay product (Appendix 5). ClustalW alignments with the expected sequences revealed that the single reads for regions A and B were obtained with 100% and 99.6% accuracy, respectively (Figures 4.23 and 4.24). Apparent polymorphisms between actual data and sequence data, which are close to the sequencing primers and in more downstream regions, are likely due to sequencing errors. Other sequencing runs would have eliminated these discrepancies but was deemed unnecessary based on the high quality of the data for the desired regions. The sequence data provides convincing evidence of correct fusion of P2 and D in fusion

product 2, and correct integration of this product at the *stzA* locus. Since neither *pyr4* fragment can form a functional marker without HR with its partner fragment, it can be safely assumed that the second *pyr4* fragment (P1; part of fusion product 1) must also have integrated at the *stzA* locus. Obviously, a functional *pyr4* gene must have been generated considering that *stzA* deletion transformants were able to grow on selective media lacking uridine and uracil.

All five stable transformants appeared to possess correct integration of the *pyr4* gene at the *stzA* locus, indicative of successful gene targeting (Figures 4.21 and 4.22). This is perhaps not surprising. In gene-targeting experiments involving sixty genes in *A. nidulans*, 90% of transformants carried correct homologous integration of the gene-targeting substrates at the target site (Nayak *et al*., 2006). Additionally, Southern analysis has shown that ectopic integration events in addition to the correct targeting event are very rare in *nkuA* deficient strains (Nayak *et al*., 2006; Kück and Hoff, 2010). Care was taken to ensure that the *stzA* gene-targeting fusion products were purified from the correct DNA bands excised from a gel. However, even transformation procedures using crude PCR products containing multiple DNA bands have been shown to give a majority of transformants with correct gene-targeting events (Nayak *et al*., 2006).

The *A. nidulans stzA* gene was evidently targetable in both the *nkuA* (*ku80*) deletion strain A1149 and G191 (*nkuA*$^{+}$). Gene locus, in addition to the activities of HR and NHEJ, strongly determines whether a gene is easy to target. Approximately 10% of the ~600 genes targeted for deletion in *A. niger* proved difficult to target, possibly because they were close to contig borders or within silenced heterochromatic DNA regions (Carvalho *et al*., 2010). The targeting of essential genes also proves problematic. The presence of only unstable transformants (balanced heterokaryons), which do not regenerate on selective medium, is a strong indication that a gene is essential (Nayak *et al*., 2006). The high stability of all five *stzA* deletion transformants indicates that *stzA* is not an essential gene under the growth conditions used. A successful outcome of the *stzA* gene deletion strategy meant that alternatives to *stzA* gene deletion, such as downregulation of gene expression by RNA interference (reviewed in Kück and Hoff, 2010), were not necessary.

The *stzA* gene deletion strains were not complemented with the *stzA* gene. Complementation would typically involve the disadvantages associated with ectopic integration of *stzA*, possibly at multiple sites in the genome. Hence, it seems reasonable to expect the actions of a cloned gene – especially a transcription

factor – to be substantially altered by its locus of integration and copy number in transformants (Krappmann, 2007). Complementation of a gene mutation with the wild-type allele is considered important in cases where the gene function is unknown and helps to verify correct gene-targeting events in gene deletion strains (Carvalho *et al*., 2010). However, in the case of *stzA*, it was readily apparent that the *stzA* gene had been deleted by observing its distinctive phenotypes of sensitivity to salt, arginine and DNA-damaging agents that have already been characterised (O'Neil *et al*., 2002). Furthermore, wild-type phenotypes have already been observed following complementation of the *sltA1* mutation with the cloned *stzA* gene in *A. nidulans* strains STC1−4 (O'Neil *et al*., 2002). However, different levels of *sltA1* complementation were observed in these strains, possibly due to ectopic and multiple integration events that can reduce the apparent level of complementation of one or more mutations (Whitehead, 1990).

Deletion of the *pyr4* gene from the *A. nidulans* genome is possible by direct repeat recombination using 5-fluoro-orotic acid (5-FOA) to achieve counter selection. However, the *pyr4* gene was not deleted since no further gene targeting was necessary. This strategy ensured that the *stzA* deletion transformants and the pDJB3-transformed control strains possessed functional *pyr4* genes so that no phenotypes observed would be due to *pyr4* deficiency.

4.4.3 Possible disadvantages associated with *ku*-deletion strains

The deletion of the *ku80* (*nkuA*) gene promotes gene targeting in *A. nidulans* and other filamentous fungi (Nayak *et al*., 2006; Kück and Hoff, 2010) and this was exploited for the deletion of *stzA* in A1149. However, since *ku80* and *stzA* are both involved in DNA repair and both are deleted in A1149 *stzA* deletants (A2, A5 and A7), confusion could arise regarding their unique roles in DNA repair processes. Analysis of Ku-deficient strains in *A. nidulans* (deletion of both *nkuA and nkuB*) revealed no growth defects or markedly increased sensitivity to the DNA-damaging agents methyl methanesulfonate (MMS), bleomycin, camptothecin or hydroxyurea (Nayak *et al*., 2006). Nonetheless, the physiological changes associated with disrupted NHEJ pathways are not well documented for filamentous fungi, but some altered phenotypes have been observed for Ku-deficient strains (reviewed in Kück and Hoff, 2010). In *N. crassa*, deletion of the *ku70* homologue *mus51* resulted in elevated sensitivity to the mutagens MMS, ethyl methanesulfonate (EMS) and the antibiotic bleomycin compared to wild-type strains (Ninomiya *et al*., 2004). *A. oryzae* exhibits hypersensitivity to both MMS and phleomycin (Takahashi *et al.*, 2006). In *A. fumigatus*, the only obvious phenotype to be observed in either Ku70- or Ku80-deficient strains was increased MMS sensitivity (Krappmann *et al*., 2006). In the yeast *C. neoformans*, *ku70* and *ku80* double mutants showed increased sensitivity to phleomycin but not to other DNA-damaging agents (Goins *et al*., 2006).

Limitations associated with Ku-deficient strains, such as vulnerability to DNA-damaging agents, can now be overcome by transiently disrupting the NHEJ pathway to avoid any detrimental and pleiotropic effects caused with a constantly inactive NHEJ pathway. A strategy has been proposed to transiently silence the NHEJ pathway in *A. nidulans* (Nielsen *et al*., 2008). This technology involves the presence of a counter-selectable marker flanked by a direct repeat composed of *nkuA* sequences, allowing the *nkuA* gene to be restored. Thus, the initial absence of the *nkuA* gene increases gene-targeting efficiencies to increase the chances of successful gene deletion, but its subsequent restoration eliminates the risk of defective NHEJ from interfering with subsequent phenotypic analyses. This approach has been adapted for *A. niger* to produce a transiently disrupted *ku70* (*kusA* strain) for this species (Carvalho *et al*., 2010).

A second strategy could be used to restore the *ku80* gene subsequent to deletion of *stzA*. Since *A. nidulans* has a sexual cycle, sexual crosses with *ku80*$^{+}$ strains should generate progeny with a functional *ku80* gene and a deleted *stzA* gene as a result of meiotic recombination. Southern blot analysis could then be used to

confirm the restoration of the *ku80* gene. For example, a *ku70* null mutation was eliminated from a *Sordaria macrospora* strain by crossing to a spore color mutant and isolating the progeny (Pöggeler and Kück, 2006). This strategy assumes that *ku* deletion strains of *A. nidulans* are not defective in meiosis.

4.4.4 Osmotic stress sensitivity of *stzA* deletion strains

The *stzA* deletion strains (A2, A5, A7, G6 and G25), along with the *sltA1* strain GO281, exhibited increased sensitivity to a range of ionic osmotic stresses (0.5 M NaCl, 0.6 M KCl, 0.2 M $MgCl_2$, 0.3 M LiCl, 0.15 M $AlCl_3$ and 0.3 M RbCl compared to the control strains (AP1, AP2, GP1 and GP2) and the wild-type strain L20 (Figure 4.25). The *sltA1* mutation in *A. nidulans* was initially identified as conferring sensitivity to Na^+ and K^+ (Spathas, 1978). An *A. nidulans stzA* deletion strain constructed by Spielvogel *et al*. (2008), during the course of the present study, exhibited sensitivity to the cations Mg^{2+}, Li^+ and Cs^+ in addition to Na^+ and K^+. Since the phenotypic profiles of cation sensitivity for *stzA* deletion mutants in the present study are similar to those obtained by (Spielvogel *et al*., 2008), for the cations tested, these findings lend credence to successful *stzA* deletion in strains A2, A5, A7, G6 and G25. Clearly, StzA has a major role to play in cation homeostasis. Furthermore, the study by Spielvogel *et al*. (2008) demonstrated reduced expression of *ena1* in the *stzA* deletion mutant compared to the wild-type strain. Ena1 is the homologue of the *S. cerevisae* P-type ATPase that pumps Na^+ out of the cell in conditions of high extracellular Na^+. Besides Na^+, ENA ATPases mediate efflux of K^+ and Li^+ (Benito *et al*., 2002).

In addition to its involvement in cation efflux, StzA may regulate genes besides *ena1* to contribute to enhanced osmotolerance. Such genes may include those involved in the synthesis of compatible solutes. Glycerol is the main compatible solute produced in response to salt stress in *A. nidulans* (Redkar *et al*., 1995). A reduced ability of *sltA1* mutants to induce polyol formation under salt stress has been demonstrated (Clement *et al*., 1999). Normal growth of *stzA* deletants (resembling the control strains on 0.5 M NaCl) was not restored by the addition of 1% of glycerol, mannitol or erythritol to MM containing 0.5 M NaCl (Figure 4.26). The results indicate that although the *A. nidulans* strains tested must be able to sense and uptake these three polyols to enable them to be used as sole carbon sources (Section 4.3.3a), they do not function effectively as compatible solutes in *stzA* deletion strains. Glycerol, mannitol and erythritol present as sole carbon sources compensated for the lack of glycerol biosynthesis in *gldB* deletion strains to restore growth

resembling the wild-type strain in conditions of salt stress (1 M NaCl; De Vries *et al*., 2003). The *stzA* deletants showed similar phenotypic responses on MM containing 1% glucose, 0.5 M NaCl and 1% glycerol compared to 0.5 M NaCl and 1% glycerol alone (Figure 4.26). In these conditions, glycerol is thought not to be subject to glucose repression (De Vries *et al*., 2003).

Growth responses of *stzA* deletants to sources of osmotic stress that were non-ionic were also analysed (Figure 4.27) The *stzA* deletants, the control strains, L20 and GO281 were all similarly sensitive to 1.2 M sucrose and 1 M and 2 M sorbitol, which impose non-ionic osmotic stresses (Hirasawa and Yokoigawa, 2001; Hirasawa *et al*., 2006). All strains were similarly resistant to 10% PEG, a known source of osmotic stress in *A. nidulans* (Shi *et al*., 2008). These results indicate that StzA is not involved in tolerance to non-ionic stresses. The *stzA* deletants were more sensitive than the controls strains in the presence of ethanol (6%), a chaotropic solute that reduces water activity and decreases electrostatic interactions, disrupting the structure and function of macromolecules. This is further evidence of a reduced ability of *stzA* deletion strains to synthesise glycerol and erythritol, two compatible solutes known to protect against ethanol-induced chaotropic stress in *A. nidulans* (Hallsworth *et al*., 2003). The addition of either 1% glycerol or erythritol (instead of glucose) to MM containing 6% ethanol did not improve the growth of any strain (results not shown).

4.4.5 DNA damage sensitivity of *stzA* deletion strains

Comparison of growth morphologies in response to DNA-damaging agents revealed enhanced sensitivity of *stzA* deletion strains to UV, the UV mimic 4NQO and the alkylating agent MNNG compared to *stzA*$^{+}$ strains (Figure 4.28). The sensitivities of the *stzA* deletion strains to these stresses were similar in intensity to those reported for GO281 (O'Neil *et al*., 2002). During photoreactivation tests, light treatment post UV irradiation brought about a small but consistent recovery in growth for *stzA* deletion strains. This partial reversal of UV sensitivity in *stzA* deletants indicates that DNA photolyase (CryA) is still functional in these strains (Bayram *et al*., 2008). Hence, it seems unlikely that StzA regulates CryA.

MNNG results in the toxic lesion O^6-methyl guanine and the formation of phosphotriesters (Wyatt and Pitman, 2006). The DNA damage response to alkylating agents involves the methylphosphotriester (MPT)

and O^6-alkylguanine DNA alkyltransferase (AGT) DNA repair proteins in both *A. nidulans* and *A. fumigatus* (O'Hanlon *et al*., 2012). Hence, these two proteins in *A. nidulans*, encoded by AN2276 and AN4587, may be under StzA control. Sensitivity to alkylating agents can be due to a cell wall or cell membrane permeability defect. This is the case for *A. nidulans* strains carrying the *sagA* (sensitive to alkylating agents) mutation and is not due to a mutation within a DNA repair pathway (Jones *et al*., 1999). SagA shares homology with *S. cerevisiae* End3, which is involved in endocytosis, organisation of the actin cytoskeleton and distribution of chitin (Raths *et al*., 1993; Bénédetti *et al*., 1994). As all *stzA* deletion strains and GO281 responded similarly to the alkylating agent MNNG, it is highly unlikely that all *stzA* deletants (derived from both A1149 and G191) would have a cell wall or cell membrane permeability defect, in addition to being sensitive to other sources of DNA damage.

UV and 4NQO cause bulky lesions that are principally repaired by nucleotide excision repair (NER; reviewed in Goldman and Kafer, 2004). Such repair requires the activity of either of the two *N. crassa* homologues *mus18* or *mus38*, encoded by AN0604 and AN8713 in *A. nidulans*. Double mutants for both of these homologues are expected to show severe sensitivity to UV irradiation (Hatakeyama *et al*., 1998; Goldman and Kafer, 2004). However, *stzA* deletants were able to grow in the presence of both UV and the UV mimic 4NQO, albeit at a reduced rate. This indicates that only one NER pathway is subject to regulation by StzA. It can be hypothesised that StzA regulates the AN8713 gene due to the presence of 9 putative StzA binding sites in its promoter compared with just 2 in the promoter of the AN0604 gene (5.2.7cii). An alternative to the NER pathway for the repair of bulky lesions is the HR pathway, requiring *uvsC* (van Heemst *et al*., 1997), which has 5 putative StzA binding sites in its promoter.

4.4.6 Carbon source utilisation of *stzA* deletion strains

It has been shown that the *sltA1* mutant GO281 had a reduced ability to utilise glycerol and ethanol as sole carbon sources (O'Neil *et al*., 2002). Results in Chapter 3 were indicative of an uncharacterised secondary mutation in GO281, affecting growth of this strain on numerous carbon sources (mainly gluconeogenic), which was unrelated to the *sltA1* mutation. Growth responses of *stzA* deletion strains were analysed on a range of carbon sources that fed into the glycolytic or gluconeogenic pathway and also plant substrates. All the glycolytic carbon sources tested: glucose, fructose, sorbitol, mannitol, erythritol and galactose supported equally strong growth of *stzA* deletants and *stzA*$^{+}$ strains (Figure 4.29). The results indicate that the glycolytic pathway is not disrupted in *stzA* deletion strains. The *stzA* deletion strains grew similarly to *stzA*$^{+}$ strains on the gluconeogenic carbon sources ethanol and the amino acids tested (glutamic acid, glutamine, proline and leucine; Figure 4.30). All strains grew well on starch and reasonably well on xylose and xylan (Figure 4.32).

Poor growth of strains originating from G191 (the *stzA* deletants G6 and G25 and the controls GP1 and GP2) on glycerol and galacturonic acid is indicative of an uncharacterised mutation in G191 affecting utilisation of these carbon sources (Figures 4.30 and 4.32). Analysis of *stzA* deletion transformants originating from A1149 indicates that StzA is not involved in the metabolism of glycerol, which is the main compatible solute produced in response to osmotic stress in *A. nidulans* (Redkar *et al*., 1995). Hence, StzA appears not to affect the actions of the key enzymes of the glycerol metabolic pathway: glycerol dehydrogenase, glycerol kinase and glycerol-3-phosphate dehydrogenase (reviewed in De Vries *et al*., 2003; Salazar *et al*., 2009). It seems that deletion mutants of both StzA and its orthologue in *T. reesei* (Ace1) are unaffected in their ability to utilise glycerol (Aro *et al*., 2003), indicating neither protein regulates glycerol metabolism.

The *stzA* deletion strains derived from both A1149 and G191 grew poorly when compared to the control strains on acetic acid, succinic acid and citric acid, compounds that are metabolised *via* the TCA cycle (Figure 4.31). Growth of the *stzA* deletion strains improved on acetic acid and succinic acid when their concentrations were reduced from 1% to 0.25%. This is thought to be due to the reduction in the Na^{+} in NaOH required for the adjustment of media to pH 6.5. However, this does not explain why *stzA* deletants still grew poorly on citric acid at a concentration of 0.25%.

The *stzA* deletants grew poorly compared to the control strains on putrescine, indicating that StzA may regulate the utilisation of putrescine (Figure 4.30). Putrescine is one of the polyamines produced by arginine metabolism that are used to stabilise various cellular components in *A. nidulans* during normal growth and development (Walters *et al*., 1997). Hence, the poor utilisation of putrescine by *stzA* deletants may be related to the known disruption of arginine metabolism in *sltA1* strains (Clement *et al*., 1996).

Collectively, the results would appear to indicate that StzA has no role to play in the regulation of gluconeogenesis. The results in this chapter are further evidence to support the notion that the disrupted carbon metabolism in GO281 is not due to the *sltA1* mutation but due to an uncharacterised secondary mutation within this strain (Chapter 3).

4.4.7 Plant polysaccharide-degrading enzyme production of *stzA* deletion strains

Aspergillus species, including *A. nidulans*, are known for their ability to produce an array of enzymes that degrade plant cell wall polysaccharides (reviewed in De Vries *et al*., 2001; Coutinho *et al*., 2009). *A. nidula*ns *stzA* deletants demonstrated reduced growth compared to *stzA*$^{+}$ strains on pectin, polygalacturonic acid and galacturonic acid when provided as sole carbon sources (Figure 4.32). Furthermore, Ace1 is a known regulator of cellulase and xylanase synthesis in *T. reesei* (Saloheimo *et al*., 2000; Aro *et al*., 2003). For these reasons, it was decided to perform the DNS assay to detect (indirectly) levels of cellulases, xylanases and polygalacturonases produced by strains A2 and A1149. (These two strains are isogenic apart from A1149 lacking *pyr4* and A2 lacking *stzA*.) DNS reacts with reducing sugars to produce 3-amino-5-nitro salicylic acid, which absorbs light strongly at 540 nm (Miller, 1959). Hence, the assay was able to detect levels of the reducing sugars glucose, xylose and galacturonic acid liberated by these enzymes in the presence of cellulose, xylan and polygalacturonic acid, respectively. The strains were induced to produce cellulases, xylanases or polygalacturonases in liquid MM cultures with 1% of cellulose, xylan or polygalacturonic acid as a sole carbon source. Results of the DNS assay indicated that the *stzA* deletion strain A2 and the *sltA*$^{+}$ strain A1149 produced similar levels of all three types of plant cell wall polysaccharide degrading enzymes after the 48 hour growth period (Tables 4.6–4.8).

4.4.8 Nitrogen source utilisation of *stzA* deletion strains

All *stzA* deletion strains grew equally well as the *stzA*$^{+}$ strains on ammonium tartrate, sodium nitrate, glutamic acid, glutamine, proline and phenylalanine when used as sole nitrogen sources (Figure 4.35) Although there were differences in carbon source utilisation between L20 and GO281, no differences in nitrogen source utilisation were detected in these strains when grown on a wide range of sole nitrogen sources (Chapter 3).

Ammonium is a highly preferential nitrogen source for *A. nidulans* and causes repression of alternative less favourable nitrogen sources *via* the nitrogen metabolite repressor AreA (reviewed in Marzluf, 1997; Wilson and Arst, 1998). Strong growth of *stzA* deletants on ammonium tartrate indicates no differences in AreA-mediated nitrogen metabolite repression in the presence of ammonium. Strong growth of *stzA* deletants on sodium nitrate indicates that the nitrogen assimilation pathway is unaffected by StzA. This pathway in *A. nidulans* involves NirA-mediated induction of genes (including *niiA* and *niaD*) required for the uptake and reduction of nitrate and nitrite, which are subsequently reduced to ammonia before conversion to glutamate and glutamine (Burger *et al*., 1991). These two amino acids were well utilised by all strains as sole nitrogen sources.

Results from this chapter and Chapter 3 indicate that amino acid utilisation is unaffected in *stzA* deletion strains, with the exception of arginine. When arginine was used as a sole nitrogen source (50 mM), the *stzA* deletion mutants displayed reduced radial growth and a distinctive red phenotype similar to that displayed by the *sltA1* mutant GO281 (Clement *et al*., 1996), whereas the *stzA*$^{+}$ strains grew well on arginine (Figure 4.35). L20 and GO281 exhibited similar growth phenotypes on all twenty amino acids used as sole nitrogen sources at concentrations of both 10 mM and 50 mM, with the exception of arginine (Chapter 3). The present study shows that the pH of the growth medium clearly has an impact on arginine sensitivity. The *stzA* deletion strains are highly sensitive to an arginine concentration of 50 mM at pH 8.0 but show no sensitivity at pH 6.5.

In fungi, arginine catabolism involves the conversion of arginine to ornithine and proline and finally to glutamate. The genes that regulate arginine catabolism in *A. nidulans* are *agaA*, *otaA, arcA* and the global carbon and nitrogen repression systems (Empel *et al*., 2001; Dzikowska *et al*., 2003; Macios *et al*., 2012).

The first step in this process is carried out by arginase, encoded *by agaA*, which catalyses the conversion of arginine to ornithine and urea (Borsuk *et al*., 1999). The arginine sensitivity phenotype is largely relieved by ornithine, the first product in the catabolism of L-arginine by arginase (Clement *et al*., 1996). Northern analysis has revealed that *agaA* is likely to be regulated directly or indirectly by StzA (Barham-Morris, 2006). Furthermore, it has been shown that reduced arginase activity in GO281 is directly linked to salt tolerance. A three-fold reduction in arginase activity in GO281 compared to L20 was observed during a 3.5-hour experimental period following two hours of 0.5 M NaCl stress treatment (Attwell *et al*., 1995; Barham-Morris, 2006). This led to the hypothesis that the guanidinium moiety of arginine competes with Na^+ for the Na^+ binding site within the Na^+/H^+ antiporter to modulate its activity. Na^+/H^+ exchanger activity was inhibited in response to increasing intracellular arginine concentrations in the presence amiloride, a known inhibitor of the Na^+/H^+ exchanger with remarkable similarity in chemical structure to arginine (Attwell, 1995; Woodcock, 1997).

4.4.9 pH sensitivity of *stzA* deletion strains

Regulation of pH in *A. nidulans* is mediated by PacC (Tilburn *et al*., 1995). Similar growth responses of *stzA* deletants at pH 5.0, 6.5 and 8.0 when compared to the *stzA*$^+$ strains indicated that StzA has no effect on pH regulation. However, it had been previously shown that an *stzA* deletion strain is highly sensitive to alkaline pH in the presence of a buffer (100 mM Na_2HPO_4; Spielvogel *et al*., 2008) and these results were reproduced in the present study. It is unknown why *stzA* deletants display a stark growth difference at pH 8.0 that is dependent on the presence/absence of the buffering agent.

The ability of *stzA* deletion strains to swell and germinate at different pH has yet to be tested. However, at acidic pH values (pH 5.0), *A. nidulans sltA1* strains demonstrated abnormal ability to swell and germinate when compared to wild-type strains (O'Mahony *et al*., 2002). This behaviour resembles the ability of *A. fumigatus* spores to germinate in the acidic environment of engulfing phagosomes, and therefore the *stzA* gene may be involved in the perception of favourable conditions to determine when a spore breaks dormancy to colonise a habitat (Waslynka and Moore, 2003).

4.4.10 Oxidative stress and fungicide responses of *stzA* deletion strains

In *A. nidulans*, responses to both osmotic and oxidative stresses are regulated by the general stress signalling HOG pathway, in which AtfA interacts with the MAPK HogA (Hagiwara *et al*., 2009; Balázs *et al*., 2010; Lara-Rojas *et al*., 2011). Oxidative stress is imposed by reactive oxygen species (ROS), cytotoxic by-products generated by aerobic respiration. ROS include hydrogen peroxide, which is toxic mainly due to the more reactive species that are derived from it, such as hydroxyl radical and singlet oxygen (reviewed in Li *et al.*, 2009; Hansberg *et al*., 2012). Detoxification of hydrogen peroxide in *A. nidulans* involves at least four differentially regulated catalase genes: *catA*–*catD* (Kawasaki and Aguirre, 2001). It seems that none of these genes is subject to regulation by StzA considering that *stzA* deletion and *stzA*$^{+}$ strains were similarly sensitive to exogenous hydrogen peroxide (Figure 4.37). These results were supported by similar growth responses of all strains to menadione, which exerts not only oxidative stress but also non-oxidative stress by chemically modifying (arylating) cell components and enhancing membrane fluidity (Shertzer *et al*., 1992). The *A. nidulans* HOG pathway is also known to be activated by the fungicide fludioxonil (Hagiwara *et al*., 2009). As *stzA* deletants and *stzA*$^{+}$ strains showed similar sensitivities to fludioxonil (Figure 4.38), it would seem that StzA does not contribute to the functioning of the HOG pathway in the presence of this fungicide.

Another pathway involved in tolerance to stress responses in *A. nidulans* is the cell wall integrity (CWI) pathway. This pathway is a major regulator of cell wall biogenesis resulting from environmental stresses that include hypo-osmolarity and heat shock (Levin, 2011). The CWI pathway is well characterised in *S. cerevisiae*, and is highly conserved in yeast and other fungal genomes (Nikolaou *et al*., 2009). In *A. nidulans*, this pathway involves the sensors WscA and WscB and the signal-transducing genes *rhoA*, *pkcA*, *mpkA* and *rlmA* (Futagami *et al*., 2011). Several components of the CWI pathway are involved in tolerance to antifungal agents in *A. nidulans*, with *rhoA*, *mpkA*, and *rlmA* deletion strains exhibiting sensitivity to calcoflour white (Guest *et al*., 2004; Fujioka *et al*., 2007). Since *sltA1*/*stzA* deletion strains exhibited increased sensitivity to both hypo-osmolarity (O'Neil *et al*., 2002) and the antifungal agent neomycin (Spielvogel *et al*., 2008; Figure 4.38), it seemed important to determine if these phenotypes were associated with a defective CWI response. Hence, it seemed appropriate to analyse the sensitivity of *stzA* deletion strains to calcofluor white and congo red, which inhibit cell wall biosynthesis, to deduce if the CWI pathway was disrupted in these strains. Calcofluor white and congo red interact with different components in the cell wall, with calcofluor white binding to nascent chitin to inhibit the assembly of chitin chains (Ram *et al*., 1994) and congo red inhibiting β-

1,3-glucan assembly (Kopecká and Gabriel, 1992). The finding that *stzA* deletants were no more sensitive than the control strains to calcofluor white and congo red indicates that the CWI pathway functions normally in *stzA* deletion strains (Figure 4.38). Furthermore, no binding sites for the key regulator of the CWI pathway, RlmA, were identified in the 2 kb promoter of the *A. nidulans stzA* gene (Fujioka *et al*., 2007; Chapter 5). Additionally, all *stzA* deletion strains were no more sensitive than *stzA*$^{+}$ strains to amphotericin B, which binds to ergosterol in cell membranes, leading to cell death (Baginski and Czub, 2009).

4.4.11 Stress responses of *T. reesei ace1* deletant and wild-type strains

It has been shown that *ace1* regulates cellulase and xylanase expression in *T. reesei* (Aro *et al*., 2003). However, it was unknown whether *T. reesei ace1* deletants exhibited increased sensitivity to the same stresses as *A. nidulans stzA* deletants when compared to wild-type *T. reesei* strains. The *T. reesei* wild-type and *ace1* deletion strains exhibited similar growth responses to the osmotic stresses NaCl and KCl; the DNA-damaging agents UV, 4NQO and MNNG; 50 mM arginine as a sole nitrogen source; neomycin; and acidic, neutral and alkaline pH (Figure 4.39). These results indicate that *T. reesei* Ace1 does not regulate the same stresses as its StzA orthologue in *A. nidulans*.

4.4.12 Conclusion

The construction of *stzA* deletion strains and subsequent phenotypic analyses allowed the phenotypes regulated by StzA to be ascertained precisely. This was necessary because the *sltA1* strains GO281, WV101 and STC1−4 all appear to possess at least one uncharacterised secondary mutation affecting carbon metabolism (Chapter 3; O'Neil *et al*., 2002; Barham-Morris, 2006). Furthermore, phenotypes associated with the *sltA1* mutation were only partially complemented by transformation of GO281 with the *stzA* gene (O'Neil *et al*., 2002). Phenotypic analyses of *stzA* deletants and *stzA*$^{+}$ strains implied that *stzA* is allelic with *sltA* and verified that StzA regulates tolerance to salt, DNA-damaging agents and high arginine concentrations when provided as a sole nitrogen source (Figures 4.25, 4.28 and 4.35). The similar severe sensitivity of GO281 and *stzA* deletants to these stresses indicates that the PTC in *sltA1* represents a total loss-of-function mutation. It was confirmed that StzA has no regulatory role in the utilisation of carbon sources (Figures 4.29–4.31).

Similar, strong growth of *stzA* deletants and *stzA*$^{+}$ strains indicated that StzA does not regulate tolerance to sources of non-ionic osmotic stress (1.2 M sucrose, 1–2 M sorbitol and 10% PEG; Figure 4.27), oxidative stress (hydrogen peroxide and menadione; Figure 4.37) or the fungicide fludioxonil (Figure 4.38) – all known to activate the *A. nidulans* HOG pathway (Hagiwara *et al*., 2009; Balázs *et al*., 2010; Lara-Rojas *et al*., 2011). Nor does StzA appear to regulate the CWI pathway, a major regulator of cell wall biogenesis resulting from environmental stresses that include hypo-osmolarity and heat shock (Table 4.38; Levin, 2011). This assertion is based on *stzA* deletants being no more sensitive than *stzA*$^{+}$ strains to congo red and calcofluor white.

A. nidulans StzA appears to have no role in cellulase or xylanase expression as revealed by the results of a DNS assay (Table 4.8). *T. reesei ace1* deletant and wild-type strains showed similar sensitivities to salt, DNA-damaging agents, arginine, neomycin, acidic and alkaline pH (Figure 4.39). These results confirm that *A. nidulans* StzA and *T. reesei* Ace1 regulate the distinct phenotypes of abiotic stress tolerance and cellulase and xylanase expression, respectively (O'Neil *et al*., 2002; Aro *et al*., 2003).

Chapter 5

Bioinformatic analyses of StzA orthologues

5.1 Introduction

5.1.1 Databases used for bioinformatic analyses

Online databases were used to obtain genes and proteins that were orthologous to *stzA* and to obtain promoter sequences for bioinformatic analyses. Basic Local Alignment Search Tool (BLAST) sequence similarity searching programs at several databases allowed DNA and protein sequence comparisons to be performed (Altschul *et al*., 1997). Various types of Blast programs exist; for example, Blastn compares a nucleotide sequence with a nucleotide database and Blastp compares an amino acid sequence with a protein database. Often a translated search gives better identification of DNA sequences because evolution conserves at the level of function (the protein level), and several codons can code for the same amino acid. The Blastp program detects regions of similarity in novel and otherwise unrelated proteins, thus allowing the functions of uncharacterised proteins to be predicted by comparison to previously characterised proteins (Frith *et al*., 2004).

Several internet resources are available that allow BLAST analyses to be performed for fungal species. The National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/) is a U.S. government-funded national resource for molecular biology information. NCBI maintains an annotated collection of nucleotide sequences and their protein translations, which are all publicly available on the GenBank database. The Broad Institute (www.broadinstitute.org/) is a genomic medicine research centre also based in the U.S. that is involved in the production and analysis of sequence data from fungal organisms that are important to basic research, medicine, agriculture and industry as part of the Fungal Genome Initiative (FGI). The Central *Aspergillus* Data REpository (C*A*DRE; http://www.cadre-genomes.org.uk/) is a resource for viewing annotated genes arising from *Aspergillus* sequencing and annotation projects (Mabey *et al*., 2004). The *Aspergillus* Genome Database (AspGD; http://www.aspergillusgenomes.org/) is a sister database of C*A*DRE and contains similar genomic sequence data and gene and protein annotations for *Aspergillus* species (Arnaud *et al*., 2010).

A limitation of the BLAST alogorithm is that it provides only pairwise alignments of DNA sequences and proteins. Multiple sequence alignments that align locally similar regions between two or more genes or proteins are potentially more useful for the discovery of functional motifs (Frith *et al*., 2004). ClustalW and

ClustalX are widely used for the multiple alignments of both nucleic acid and protein sequences and for preparing phylogenetic trees (Chenna *et al*., 2003). TCoffee uses both global and pairwise alignments and combines these with structural data to improve accuracy (Notredame *et al*., 2000).

5.1.2 Annotation of *stzA* genes

In silico gene prediction methods involve several strategies (reviewed in Sleator, 2010; Haas *et al*., 2011). All rely on homologies within databases to identify highly conserved exons (e.g. using genomic DNA, protein sequences and EST sequences from one or more genomes). Gene prediction strategies also take into account the characteristics of the DNA sequence itself to determine if the sequence encodes a protein. These characteristics include the functional sites of the introns of a gene (i.e. donor sites, acceptor sites and branchpoint motifs) and other indicators that include hexamer frequency, nucleotide composition, codon usage and base occurrence periodicity.

Automated predictions of genes often lead to incorrect annotations. Introns are often missed or incorrectly predicted using automated annotations alone due to incorrect assignment of intron/exon boundaries or the merging of neighbouring loci (Brachat *et al*., 2003; Wortman *et al*., 2009). Even in the well-characterised yeast *S. cerevisiae*, almost 30% of intron predictions were found to be incorrect (Davis *et al*., 2000b). Rep e*t al.* (2006) estimated that even under ideal circumstances, the intron–exon structure of at least 1 in 10 genes would be incorrectly predicted using the gene annotation programs currently available. Proteome comparisons among the genomes of *A. nidulans*, *A*. *fumigatus* and *A. oryzae* revealed the majority of identified orthologue groups (~80%) contained members differing in length and/or numbers of exons due to major differences in gene annotation among species (Wortman *et al*., 2009). Inconsistencies in gene annotation were attributed to the diverse annotation processes employed at different sequencing centres and the lengthy timeframe over which the different genomes were sequenced.

Since the vast majority of protein-coding genes within sequenced fungal genomes have not been experimentally characterised, accurate methods for gene prediction are essential and should include manual annotations by constructing alignments and considering fungal gene structure (Sims *et al*., 2004). Hence, StzA sequences were examined for potential misannotations. More accurate manual annotations were

carried out using protein and DNA alignments with considerations to fungal intron structure. A variety of approaches were used to ensure correct manual annotation of StzA proteins which included:

- Aligning StzA proteins to identify any amino acid sequence extensions or omissions within or surrounding known intron positions.
- Considering intron positional profiles of StzA within the 4 classes of the Pezizomycotina subphylum.
- Considering intron length conservation.
- Identifying the fungal 3'-GU… and …AG-5' intron splice sites that correspond to known conserved intron positions in an *stzA* genomic DNA alignment (Kupfer *et al*., 2004).
- Identifiying the fungal branchpoint motif (5'-RCURAY) towards the 3' end of predicted introns (Kupfer *et al*., 2004).
- Verifying that the correct reading frame was maintained following the removal of an intron by observing expected amino acid sequence conservations in the protein region immediately downstream of the predicted intron.

5.1.3 Intron recognition

Gene expression in fungi, like that of other eukaryotes, depends on accurate splicing *via* the co-ordinated efforts of a spliceosome (reviewed in Roy and Gilbert, 2006). In metazoa, spliceosomes are known to be composed of 5 small nuclear RNAs and over 60 proteins that function as splicing factors. The spliceosome co-ordinates with conserved *cis* elements in the intron to correctly identify its boundaries. These *cis* elements comprise: 3' and 5' dinucleotide donor and acceptor splices sites (universally known to be GU and AG); a branchpoint and surrounding motif; and polypyrimidine tracts (Kupfer *et al*., 2004). Splicing complexes bind to these elements to facilitate the correct excision of intron sequences and the joining of exon sequences so that the mRNA can be translated into a functional protein

Compared to higher eukaryotes, fungi are known to possess relatively long exons and short introns that are biased towards the 5' end of genes (Nielsen *et al*., 2004). A comprehensive examination of fungal introns revealed that *A. nidulans* and *N. crassa* had narrow ranges of intron lengths with a dominant peak distribution between 50 and 70 nucleotides, whereas yeasts had much broader intron length ranges

compared to filamentous fungi (Kupfer *et al*., 2004). Greater than 98% of the fungal introns in the study belonged to the GU-AG canonical class. The vast majority of the non-canonical introns were estimated to possess GC-AG splice sites and the percentages of *A. nidulans* and *N. crassa* introns belonging to this class were determined to be 1.15% and 0.86%, respectively, compared to 1.19% for *S. cerevisiae*. These results reflect those of a similar detailed study that determined the frequency of non-canonical GC-AC introns in *N. crassa* and *Fusarium graminearum* to be 1.2% and 1.0%, respectively (Rep *et al*., 2006).

General fungal consensus sequences have been derived for the 5' splice site, 3' splice site and branchpoint motifs: GURWGU, YAG and RCURAY, respectively, where the underlined A is the branchpoint (Rep *et al*., 2006). The mean distance of the branchpoint lay between 75 and 87% of the total intron length (i.e. The branchpoint lay between nucleotides 75 and 87 for an intron length of 100 nucleotides.). Most introns (between 83.2 and 93.7%) had polypyrimidine tracts located between the 5' splice site and the branchpoint, but this did not preclude the presence of polypyrimidine tracts in the 3' region.

5.1.4 Zinc finger proteins and nuclear localisation signals

Eukaryotic transcription factors (TFs) function by recognition of specific DNA sequences and are characterised on the basis of their DNA-binding domains (Laity *et al*., 2001). The C_2H_2 zinc finger is one of the most common DNA-binding motifs in eukaryotes (Iuchi, 2001; Larabee *et al*., 2005). Seventy-two C_2H_2 zinc finger proteins have been identified in *A. nidulans*, including StzA (O'Neil *et al*., 2002; Wortman *et al*., 2009). In this protein, three C_2H_2 zinc fingers constitute the DNA-binding domain, a region that is conserved in the StzA/Ace1 proteins of other filamentous ascomycetes. The structure of a single C_2H_2 zinc finger is detailed in Section 1.4. Zinc finger proteins have diverse roles in cellular processes, including transcription and translation, DNA replication and repair, metabolism, cell signalling and proliferation, and apoptosis (Iuchi, 2001).

Nuclear localisation signals (NLSs) comprise a mechanism that initiates the import of specific proteins into the nucleus to allow TFs, such as StzA, to mediate transcription. The NLS is recognised in the cytoplasm by a heterodimeric receptor, containing an α and β subunit (Nigg, 1997). The α subunit is involved in the specific recognition of the NLS in the target nuclear protein and the β subunit docks the α subunit at the

cytoplasmic face of the nuclear pore complex. The presence of Ran-GTP in the nucleus promotes the dissociation of the α-β heterodimer and the subsequent translocation of the transported protein through the channel of the nuclear pore complex into the nucleus. The essential elements of a bipartite NLS are two clusters of basic amino acids composed of lysine (K) and arginine (R): a smaller cluster of two basic amino acids and a larger cluster of at least three basic amino acids in a group of five, separated by a spacer sequence (Dingwall and Laskey, 1998). This spacer sequence must be at least ten amino acids long to promote nuclear targeting, but its precise sequence appears not to be significant.

5.1.5 Transcriptional regulation in filamentous fungi

The phenotypic differences observed for StzA orthologues are likely to be attributed to both changes in StzA protein sequences and changes in *stzA* gene expression. The binding of TFs to sequence-specific sites in the promoters of genes is the best-characterised mechanism involved in the control of gene expression (Tirosh *et al*., 2008). TF binding sites can be difficult to identify because they are usually short (5–15 bp), degenerate and occur frequently in the genome (Tsai *et al*., 2006). Comparisons of promoter sequences from different species can reveal the identities of regulatory sequences since TF binding sites of functional significance evolve at a much slower rate compared to their neutral surrounding sequences, and therefore can be identified due to their greater levels of conservation (Wittkopp, 2010).

Comparative genomics has been used to discover novel regulatory motifs in the filamentous ascomycete *F. graminearum* by comparing four *Fusarium* genomes (Kumar *et al*., 2010). This approach resulted in the prediction of 73 TF binding motifs. Many of these are considered likely to be true functional elements since 22% showed strong sequence similarity to known binding site motifs in *S. cerevisae*, and a further 30% showed strong enrichment in functionally related gene clusters. Reliable predictions of functional TF binding sites using sequence conservation does not preclude the possibility of unconserved binding site sequences from being functional (Doniger *et al*., 2007).

Many TF binding sites have been determined experimentally for *Aspergillus* species (reviewed in Machida and Gomi, 2010). In the present study, the promoters of *stzA* orthologues were examined for the presence and positional conservation of characterised TF binding sequences. The StzA/Ace1 binding site has been

determined *in vitro* to be 5'-AGGCA in both *A. nidulans* (Spielvogel *et al*., 2008) and *T. reesei* (Aro *et al*., 2003). The phenotypic analyses presented in chapters 3 and 4 have provided insights into genes likely to be regulated by StzA. In this chapter, promoters of candidate StzA-regulated genes were analysed for the presence of potential StzA binding sites in *A. nidulans*. These included genes involved in osmotic stress, DNA repair, pH regulation and cation homeostasis, and nitrogen metabolism.

5.1.5a Osmotic stress

The *A. nidulans* genome contains gene orthologues of the HOG response MAPK pathway of *S. cerevisiae* (Section 1.3.2), which is involved in responses to osmotic stress, oxidative stress and cell cycle regulation (Furukawa *et al*., 2005). In *A. nidulans*, this pathway is also involved in stress signalling, principally osmoadaptation, but also responds to oxidative stress and fungicides (Hagiwara *et al*., 2009; Balázs *et al*., 2010; Lara-Rojas *et al*., 2011). A model for His–Asp phosphorelay signalling and HogA MAPK cascade activation *via* the SskA–HogA–AtfA pathway has been proposed for *A. nidulans* (Section 1.3.5; Hagiwara *et al*., 2009). The conservation of major components of the HOG pathway in all other sequenced *Aspergillus* species indicates similar regulatory stress signalling processes among *Aspergilli* (Miskei *et al*., 2009).

Despite the similarity between yeast and *Aspergillus* HOG pathways, there is a notable absence in *Aspergillus* species of important HOG pathway transcriptional regulators (including Sko1p, Hot1p and Msn1p) and low homologies of the *Aspergillus* proteins YpdA, MsnA, AtfA, NapA, HsfA and HacA compared to their counterparts in *S. cerevisiae*. Hence, stress response mechanisms are likely to be more similar within the *Aspergillus* genus than between any one *Aspergillus* species and *S. cerevisiae* (Miskei *et al*., 2009).

5.1.5b DNA repair

A. nidulans is a model genetic system for the study of the DNA damage response in eukaryotes and serves as a protective mechanism that allows the integrity of the genome to be preserved (Goldman and Kafer, 2004). DNA damage has many causes, which include UV irradiation, X-rays and alkylating agents that can lead to single-strand breaks, double-strand breaks, base damage and DNA-protein crosslinks (Goldman *et al*., 2002). Accumulated mutations can lead to genomic instability. Hence, fungi, like other eukaryotes, have multiple DNA repair mechanisms, which include photoreactivation, nucleotide excision repair, postreplication repair, recombination repair, cell cycle arrest and apoptosis (Goldman *et al*., 2002; Semighini *et al*., 2005). These processes are described in more detail in Section 1.3.4. An analysis of the *A. nidulans* transcriptome when exposed to camptothecin-induced DNA damage revealed that many DNA repair pathways act in parallel in response to DNA damage in this organism (Malavazi *et al*., 2006).

Two main signal transduction pathways respond to DNA damage in all organisms: the ATM (mutated in ataxia telangiectasia) and the ATR (ATM-Rad3-related) pathways (Abraham, 2001). Both pathways interfere with the cell cycle and repair double-strand breaks, albeit the ATR pathway exhibits a slower response. The ATR pathway additionally responds to agents that interfere with the function of replication forks. These agents include hydroxyurea, UV light and DNA alkylating agents. In *N. crassa*, both *mus9* and *mus21* (homologues of the mammalian DNA damage checkpoint: ATR and ATM) are involved in the DNA damage response, and at least one of the pathways must be functional for survival (Wakabayashi *et al*., 2010).

5.1.5c pH regulation and cation homeostasis

Fungi possess several signalling pathways that allow them to detect and respond to changes in pH. However, adaptation to environmental pH is controlled principally by the Rim101/PacC signal transduction pathway that is conserved among filamentous ascomycetes (reviewed in Arst and Peñalva, 2003; 2004; Peñalva *et al*., 2008). The pH regulatory system in *A. nidulans* involves proteolytic processing of the PacC TF, mediated by a signal transduction pathway involving the *pal A*, *B*, *C*, *F*, *H* and *I* genes. PacC contains three C_2H_2 zinc fingers that recognise the DNA consensus sequence 5'-GCCARG (Espeso *et al*., 1997). The zinc finger region of PacC is conserved in a wide range of fungal species that include not only ascomycetes

but also basidiomycetes (Arst and Peñalva, 2004). In *A. nidulans*, PacC acts positively on genes expressed in alkaline conditions and negatively on genes expressed in acid conditions (Tilburn *et al*., 1995).

Intracellular pH is controlled additionally by K^+/H^+ and Na^+/H^+ exchangers, Ca^{2+}/H^+ exchangers and P-type H^+ and Ca^{2+} ATPases, which also serve to regulate osmolarity and calcium signalling (Benčina *et al*., 2009; Harris *et al*., 2009). The function of these transporters is to reduce cations in the cytoplasm to a resting level by transporting them across membranes out of the cell or into storage organelles (Benčina *et al*., 2009).

5.1.5d Nitrogen metabolism

The transcriptional activator AreA is responsible for the global regulation of genes required for the utilisation of nitrogen compounds in *A. nidulans* (reviewed in Marzluf, 1997; Krappmann and Braus, 2005). AreA is one of many genes involved in the nitrate assimilation pathway. In this pathway, nitrate is completely reduced to yield ammonium, which is incorporated into the amino acids glutamate and glutamine. The NirA gene product directly induces the genes required for the uptake and reduction of nitrate, with CrnA permease mediating the uptake of nitrate, and the enzymes NiaD and NiiA reducing nitrate to nitrite and ammonium (Marzluf, 1997).

Another system that plays a key role in adapting fungi to environmental stress (and in particular amino acid deprivation) is the cross pathway control system that has been extensively studied in *S. cerevisiae* (Hinnebusch and Natarajan, 2002). The master gene responsible for this control (*cpcA*) has been cloned from *A. nidulans* (Hoffmann *et al.,* 2001) and *A. fumigatus* (Krappmann *et al*., 2004). Both orthologues control a regulatory network that activates amino acid biosynthesis when the concentrations of a single amino acid are limiting. CpcB is homologous to mammalian RACK1 (receptor for activated C-kinase 1) involved in many cellular signalling processes (McCahill *et al*., 2002). *cpcA* and *cpcB* play anatagonistic roles in both cross-pathway control and sexual development. *cpcA* activates amino acid biosynthesis under conditions of amino acid deprivation, whereas *cpcB* represses the cross-pathway control network when amino acids are present. Overexpression of *cpcA* blocks sexual development by mimicking amino acid deprivation in the presence of amino acids, and the same effect occurs by deletion of *cpcB* (Hoffmann *et al*., 2000).

5.1.6 LTR retrotransposons

A sequence associated with an LTR retrotransposon has been identified in the promoters of 11 *stzA* genes (Chilton *et al*., 2008). This sequence, 5'-CTATCAGGCA, consisting of overlapping putative AreA/StzA binding sites (Wilson and Arst, 1998; O'Neil *et al*., 2002), is found within the *pol* regions of the *REAL* LTR retrotransposon of *Alternaria alternata* and other filamentous fungi (Kaneko *et al*., 2000; Daboussi and Capy, 2003). Hence, the 5'-CTATCAGGCA sequence within the promoter of *stzA* orthologues was subsequently named *REALALE* (*REAL*-associated little element; Chilton *et al*., 2008). Virtually all present-day eukaryotes contain mobile elements consisting of moderately repetitive DNA sequences that are able to migrate within the genome (Daboussi and Capy, 2003). Long terminal repeat (LTR) retrotransposons are mobile elements whose transposition is mediated by reverse transcription. They possess DNA sequences resembling retroviruses with the distinguishing feature of long terminal repeats (Novikova *et al*., 2007).

5.1.7 Aims

The overall aim was to clarify StzA function by identifying and characterising StzA orthologues using bioinformatics analyses.

5.2 Results

5.2.1 Identification of StzA orthologues

Genes and protein orthologues of StzA were identified using Blastn and Blastp searches, respectively, at NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi; Tables 5.1 and 5.2). Much of the relevant data at NCBI originates from fungal genome projects that include several *Aspergillus* species (Galagan *et al*., 2005). All proteins with e-values of approximately $3e^{-30}$ or less were taken as orthologues of StzA. Since *stzA* encodes three C_2H_2 zinc fingers (O'Neil *et al*., 2002), other matches were found to this region (normally $3e^{-8}$ or greater) but were not considered to be functional StzA orthologues unless similarity extended into the N- and C-terminal regions.

Twenty-nine fungal species were found to contain an StzA orthologue. It was determined that StzA is specific to the Ascomycota and restricted to the Pezizomycotina subphylum. Within the Pezizomycotina, only four fungal classes were represented: Eurotiomycetes, Sordariomycetes, Dothidiomycetes and Leotiomycetes (Table 5.3). No StzA orthologues were found for the species *S. cerevisiae*, *Schizosaccharomyces pombe*, *Candida guilliermondii, C. albicans* and *Ashbya gossypii,* which belong to the Saccharomycotina and Taphrinomycotina subphyla. Nor were orthologues found in the microsporidia (*Encephalitozoon cuniculi*), the recently proposed (Hibbett *et al.,* 2007) subphylum *incertae sedis Mucoromycotina* (*Rhizopus oryzae*) or the Basidiomycota (*Ustilago maydis* and *Phanerochaete chrysosporium*).

Table 5.1. A summary of Blastn hits for *stzA* gene orthologues. *A. nidulans* genomic *stzA* was used as the query sequence.

Species (strain)	Accession number	Predicted gene function	Size (bp)	Scores (bits)	Total Score	E value	Identities
Aspergillus nidulans (FGSC A4)	XM_655431 GeneID: 2874155 AN2919	Hypothetical protein partial mRNA	2097	3570 bits (3958)	3570	0.0	2096/2202 (95%)
Neosartorya fischeri (NRRL 181)	XM_001263638 GeneID: 4590285 NFIA_069130	C_2H_2 type zinc finger domain protein partial mRNA	2538	253 bits (280) 55.4 bits (60)	309	2e-63 0.001	367/512 (71%) 93/135 (68%)
Aspergillus clavatus (NRRL1)	XM_001270816 GeneID: 4702851 ACLA_035940	C_2H_2 type zinc finger domain protein partial mRNA	2454	226 bits (250) 69.8 bits (76)	296	3e-55 5e-08	350/491 (71%) 73/95 (76%)
Aspergillus fumigatus (Af293)	XM_749720 GeneID: 3512396 AFUA_3G08010	C_2H_2 TF (Ace1), putative partial mRNA	2667	221 bits (244) 57.2 bits (62)	278	1e-53 3e-04	347/493 (70%) 82/116 (70%)
Aspergillus flavus (NRRL3357)	XM_002373607.1 Gene ID: 7917538 AFLA_087350	C_2H_2 TF (Ace1), putative	2652	113 bits (124) 100 bits (110) 57.2 bits (62)	270	8e-22 5e-18 5e-05	191/277 (69%) 177/254 (70%) 52/66 (79%)
Aspergillus oryzae (RIB40)	XM_001818368 GeneID: 5990365 AO090005001502	Hypothetical protein partial mRNA	2226	210 bits (232) 57.2 bits (62)	267	2e-50 3e-04	368/532 (69%) 52/66 (78%)
Aspergillus niger (CBS 513.88)	XM_001397480 GeneID: 4988597 An16g02040	Hypothetical protein partial mRNA	2220	181 bits (200)	181	1e-41	333/487 (68%)
Gibberella zeae	AY504867	Zinc finger TF ACE1 (*ace1*)	3111	179 bits (198)	179	4e-41	266/370 (71%)
Nectria haematococca (mpVI 77-13-4)	XM_003053497 GeneID: 9664418 NECHADRAFT_57385	Hypothetical protein mRNA	2644	172 bits (190)	172	1e-38	292/410 (71%)
Coccidioides immitis (RS)	XM_001248578 GeneID: 4568003 CIMG_02350	Zinc finger TF partial mRNA	2370	158 bits (174)	158	1e-34	273/397 (68%)
Uncinocarpus reesii (1704)	AAIW01000091	Hypothetical protein partial mRNA	1920	87.8 bits (96) 55.4 bits (60)	143	3e-14 2e-04	162/238 (69%) 81/115 (71%)
Podospora anserina (DSM 980)	XM_001908878 GeneID: 6193116 PODANSg5948	Hypothetical protein partial mRNA	2223	143 bits (158)	143	3e-30	253/366 (69%)
Ajellomyces capsulatus (NAm1)	XM_001543143 GeneID: 5451138 HCAG_00239	Hypothetical protein similar to Ace1	2760	141 bits (156)	141	1e-29	269/395 (68%)
Trichophyton rubrum (CBS 118892)	XM_003237291 Gene ID: 10378717 TERG_02061	Hypothetical protein mRNA	2487	132 bits (146)	132	9e-27	157/211 (74%)
Paracoccidioides brasiliensis (Pb01)	XM_002796661 Gene ID: 9099672 PAAG_01715	Zinc finger TF *ace1* mRNA	3120	132 bits (146)	132	9e-27	235/343 (69%)
Botryotinia fuckeliana (B05.10)	XM_001549468 GeneID: 5430014 BC1G_11939	Hypothetical protein partial mRNA	2169	129 bits (142)	129	7e-26	200/286 (69%)
Arthroderma benhamiae (CBS 112371)	XM_003015408 Gene ID: 9521181 ARB_06580	Putative C_2H_2 TF (Ace1) mRNA	2550	127 bits (140)	127	4e-25	156/211 (74%)
Hypocrea jecorina (VTT-D-80133)	AF190793	C_2H_2 zinc finger TF ACEI (*ace1*) gene	4894	125 bits (138)	125	8e-25	255/375 (68%)
Chaetomium globosum (CBS 148.51)	XM_001219996 GeneID: 4386875 CHGG_00776	Hypothetical protein partial mRNA	2046	122 bits (134)	122	1e-23	256/371 (69%)
Talaromyces emersonii	AY072919	Zinc finger TF ACE I gene, complete cds	3723	114 bits (126) 89.7 bits (98)	204	1e-21 6e-14	167/236 (70%) 132/183 (72%)
Arthroderma gypseum (CBS 118893)	XM_003172527 Gene ID: 10027848 MGYG_05167	Zinc finger TF *ace1*	2511	116 bits (128)	116	7e-22	162/227 (71%)
Verticillium albo-atrum (VaMs.102)	XM_003003504 Gene ID: 9536350 VDBG_05992	Zinc finger TF *ace1*	1752	105 bits (116)	105	1e-18	181/260 (70%)
Sordaria macrospora (k-hell)	XM_003346908 Gene ID: 10804370 SMAC_08482	Hypothetical protein mRNA	2325	102 bits (112)	102	1e-17	169/242 (70%)
Neurospora crassa (OR74A)	XM_958834 GeneID: 3880076 NCU09333	Hypothetical protein partial mRNA	2145	100 bits (110)	100	3e-17	183/267 (68%)
Penicillium chrysogenum (Wisconsin 54-1255)	AM920435	Complete genome, contig Pc00c20	2746	93.3 bits (102) 84.2 bits (92)	177	5e-15 2e-12	89/114 (78%) 167/244 (68%)
Magnaporthe grisea (70-15)	XM_361983 GeneID: 2678215 MGG_04428	Hypothetical protein partial mRNA	2214	78.8 bits (86)	78.8	1e-10	229/348 (65%)
Pyrenophora tritici-repentis (Pt-1C-BFP)	XM_001938228 GeneID: 6346206 PTRG_07931	Zinc finger TF *ace1* mRNA	2633	71.6 bits (78)	71.6	2e-08	174/261 (66%)
Phaeosphaeria nodorum (SN15)	XM_001805354 GeneID: 5982334 SNOG_15249	Hypothetical protein partial mRNA	3053	69.8 bits (76)	69.8	5e-08	53/63 (84%)
Aspergillus terreus (NIH2624)	XM_001209054 GeneID: 4315993 ATEG_01689	Predicted protein partial mRNA	2079	62.6 bits (68)	62.6	8e-06	147/220 (66%)

Table 5.2. A summary of Blastp hits for StzA protein orthologues. The *A. nidulans* StzA protein was used as the query sequence.

Species (strain)	Accession number	Predicted protein function	Size (aa)	Score (bits)	E value	Identities	Positives
Aspergillus nidulans (FGSC A4)	XP_660523	Hypothetical protein	698	1457 bits (3771)	0.0	698/698 (100%)	698/698 (100%)
Aspergillus flavus (NRRL3357)	EED58036	C_2H_2 TF (Ace1), putative	724	526 bits (1355)	3e-147	327/744 (43%)	444/744 (59%)
Aspergillus clavatus (NRRL1)	XP_001270817	C_2H_2 type zinc finger domain protein	817	521 bits (1342)	9e-146	290/570 (50%)	381/570 (66%)
Aspergillus oryzae (RIB40)	XP_001818420	Hypothetical protein	741	520 bits (1340)	2e-145	327/761 (42%)	441/761 (57%)
Neosartorya fischeri (NRRL181)	XP_001263639	C_2H_2 type zinc finger domain protein	845	515 bits (1327)	4e-144	323/740 (43%)	418/740 (56%)
Talaromyces emersonii	AAL69549	Zinc finger TF Ace1	746	507 bits (1305)	2e-141	290/592 (48%)	380/592 (64%)
Aspergillus niger (An16g02040)	XP_001397517	Hypothetical protein	739	506 bits (1304)	2e-141	295/694 (42%)	404/694 (58%)
Aspergillus fumigatus (Af293)	XP_754813	C_2H_2 TF (Ace1)	888	497 bits (1280)	1e-138	323/783 (41%)	420/783 (53%)
Penicillium chrysogenum (Wisconsin 54-1255)	CAP85969	Strong similarity to Cys_2-His_2 zinc finger TF Ace1 - *Hypocrea jecorina*	824	446 bits (1148)	2e-123	267/568 (47%)	357/568 (62%)
Coccidioides immitis (RS)	XP_001248579	Zinc finger TF	789	421 bits (1082)	1e-115	255/550 (46%)	326/550 (59%)
Paracoccidioides brasiliensis (PAAG_01715)	XP_002796707	Zinc finger TF Ace1	889	407 bits (1045)	1e-127	258/679 (38%)	367/679 (54%)
Ajellomyces capsulatus (NAm1)	XP_001543193	Hypothetical protein	919	387 bits (993)	2e-105	257/646 (39%)	354/646 (54%)
Arthroderma gypseum (CBS 118893)	XP_003172575	Zinc finger TF Ace1	836	381 bits (978)	2e-118	263/685 (38%)	377/685 (55%)
Arthroderma benhamiae (CBS 112371)	XP_003015454	C_2H_2 TF (Ace1)	849	362 bits (930)	3e-111	268/716 (37%)	386/716 (54%)
Aspergillus terreus (NIH2624)	XP_001209054	Predicted protein	692	359 bits (922)	4e-97	223/556 (40%)	314/556 (56%)
Uncinocarpus reesii (1704)	XP_002541887	Zinc finger TF Ace1	656	357 bits (915)	4e-97	209/493 (43%)	282/493 (58%)
Trichophyton rubrum (CBS 118892)	XP_003237339	Hypothetical protein	828	357 bits (917)	1e-109	262/710 (37%),	380/710 (54%)
Nectria haematococca (mpVI 77-13-4)	XP_003053543	Hypothetical protein	742	345 bits (884)	1e-105	238/644 (37%)	343/644 (53%)
Hypocrea jecorina (QM9414)	Q9P8W3	Zinc finger TF Ace1	733	335 bits (859)	8e-90	213/505 (42%)	292/505 (57%)
Gibberella zeae	AAT35113	Zinc finger TF Ace1	741	329 bits (844)	4e-88	225/587 (38%)	312/587 (53%)
Magnaporthe grisea (70-15)	XP_361983	Hypothetical protein	737	316 bits (810)	4e-84	215/554 (38%)	304/554 (54%)
Podospora anserina (DSM980)	XP_001908913	Unnamed protein product	740	316 bits (810)	5e-84	206/532 (38%)	298/532 (56%)
Chaetomium globosum (CBS 148.51)	XP_001219997	Hypothetical protein	681	313 bits (801)	4e-83	226/600 (37%)	315/600 (52%)
Phaeosphaeria nodorum (SN15)	XP_001805406	Hypothetical protein	767	311 bits (796)	2e-82	236/662 (35%)	338/662 (51%)
Sordaria macrospora (SMAC_08482)	XP_003346956	Hypothetical protein	774	303 bits (777)	1e-89	212/535 (40%)	290/535 (54%)
Verticillium albo-atrum (VDBG_05992)	XP_003003550	Zinc finger TF Ace1	583	297 bits (761)	2e-89	203/524 (39%)	279/524 (53%)
Botryotinia fuckeliana (B05.10)	XP_001549518	Hypothetical protein	722	294 bits (753)	2e-77	192/484 (39%)	273/484 (56%)
Neurospora crassa (OR74A)	XP_963927	Hypothetical protein	714	289 bits (740)	6e-76	206/498 (41%)	283/498 (56%)
Pyrenophora tritici-repentis (Pt-1C-BFP)	XP_001938263	Zinc finger TF Ace1	750	274 bits (701)	2e-71	202/594 (34%)	293/594 (49%)

Table 5.3. Classes and orders of species containing StzA orthologues. All species containing StzA orthologues were restricted to four classes in the Pezizomycotina subphylum of the Ascomycota. **Magnaporthe grisea* has not been designated to an existing order but many recent studies suggest it is more closely related to the Sordariales (Wang *et al*., 2009). Alternative names for species are shown in brackets and refer to anamorphs (asexual reproductive stage) as opposed to teleomorphs (sexual reproductive stage).

Class	Order	Species
Eurotiomycetes	Eurotiales	*Aspergillus nidulans* *Aspergillus fumigatus* *Neosartorya fischeri* (*Aspergillus fischerianus*) *Aspergillus clavatus* *Aspergillus oryzae* *Aspergillus flavus* *Aspergillus niger* *Aspergillus terreus* *Talaromyces emersonii* *Penicillium chrysogenum*
	Onygenales	*Uncinocarpus reesii* *Ajellomyces capsulatus* (*Histoplasma capsulatum*) *Coccidioides immitis* *Arthroderma benhamiae* *Arthroderma gypseum* (*Microsporum gypseum*) *Trichophyton rubrum* *Paracoccidioides brasiliensis*
Dothideomycetes	Pleosporales	*Phaeosphaeria nodorum* (*Stagonospora nodorum*) *Pyrenophora tritici-repentis*
Sordariomycetes	Sordariales	*Chaetomium globosum* *Neurospora crassa* *Podospora anserina* *Sordaria macrospora* *Magnaporthe grisea**
	Hypocreales	*Hypocrea jecorina* (*Trichoderma reesei*) *Gibberella zeae* (*Fusarium graminearum*) *Nectria haematococca* *Verticillium albo-atrum*
Leotiomycetes	Helotiales	*Botryotinia fuckeliana* (*Botrytis cinerea*)

5.2.2 Identification of introns

All of the DNA and protein sequences obtained were unannotated (i.e. the introns remained unpredicted) or had previously been annotated solely by computer analysis, with two exceptions: *stzA* from *A. nidulans* and *ace1* from *T. reesei* (*H. jecorina*). *stzA* was annotated using genomic DNA EST comparisons (O'Neil *et al*., 2002) and *ace1* using cDNA genomic sequence comparisons (Saloheimo *et al.,* 2000). The *Talaromyces emersonii* and *Gibberella zeae* sequences had been submitted as individual accessions on NCBI (i.e. not part of DNA sequencing projects). It became apparent that StzA orthologues had introns present in one or more of four conserved regions in their N-termini depending on the class to which they belonged (Figure 5.1): Dothidiomycetes (positions 1, 2 and 3); Leotiomycetes (positions 2 and 3); Sordariomycetes (positions 2 and 3), except *M. grisea* and *V. albo-atrum* (position 2); and Eurotiomycetes (multiple intron profiles, but all species contained introns 1 and 2, except *A. nidulans* lacked intron 2). These apparent conservations of intron position served to aid in the correct identification of introns. Information regarding fungal intron structure was used to correctly assign missed or misannotated introns. Fungal consensus intron sequences have previously been derived for the 5' splice site, 3' splice site and the branchpoint sequence, which are GURWGU, YAG and RCURAY, respectively, where the underlined A is the branchpoint (Kupfer *et al*., 2004).

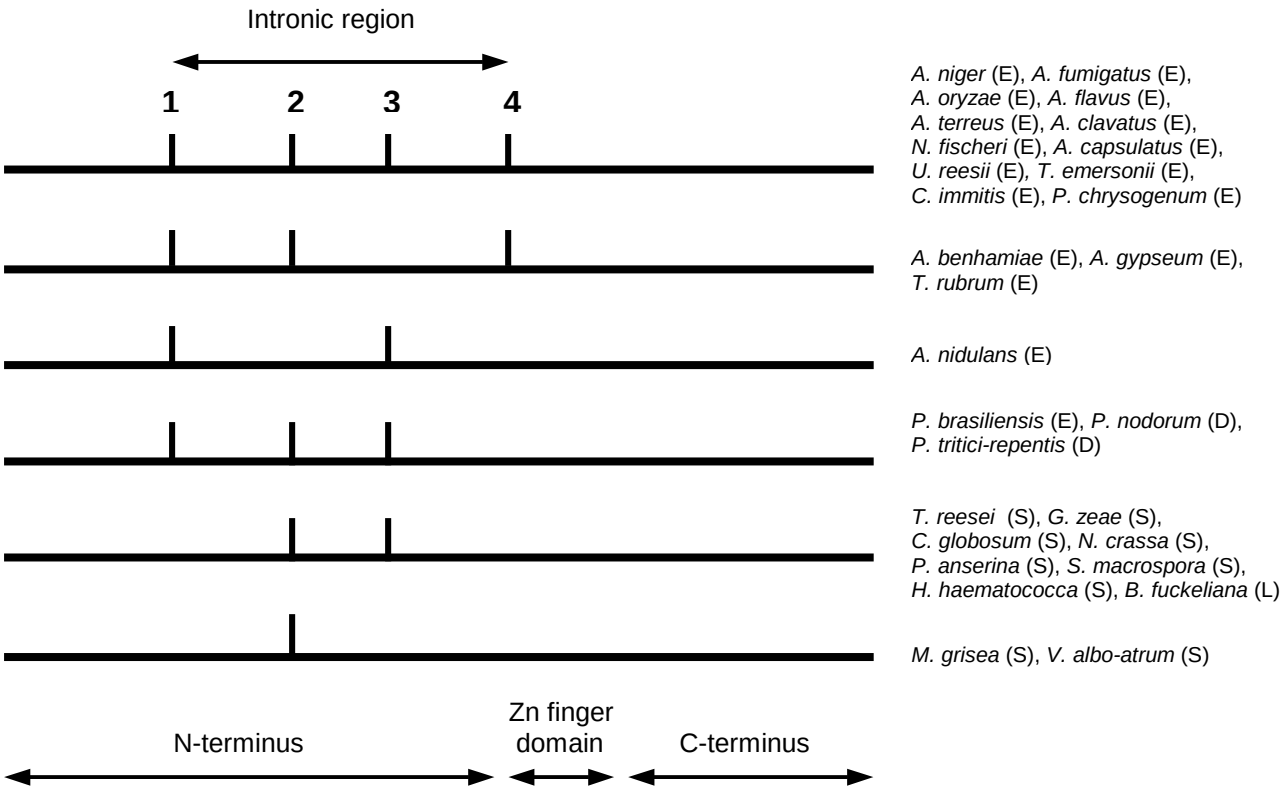


Figure 5.1. A diagrammatic representation of the relative conservation of intron positions in *stzA* orthologues. Introns could be localised to four conserved positions within the N-termini of *stzA* genes. The fungal class of the Ascomycota (Pezizomycotina subphylum) to which each species belongs is shown in brackets: Eurotiomycetes (E), Sordariomycetes (S), Dothidiomycetes (D) and Leotiomycetes (L).

Protein alignments revealed that many *stzA* genes/StzA proteins had been misannotated at NCBI (http://www.ncbi.nlm.nih.gov/; July 2008–October 2012), largely due to missed introns (Table 5.4). The missed introns could be identified by aligning proteins using ClustalW (Chenna *et al*., 2003) and TCoffee (Notredame *et al*., 2000), where they appeared as amino acid extensions within known conserved intronic regions. For example, amino acid extensions within the intronic regions (1, 2 and 3) of *A. fumigatus* StzA revealed the presence of these three missed introns, which could be verified by the presence of potential 3' and 5' splice sites within the corresponding positions of the *stzA* genomic DNA for this species. Javascript DNA translator (http://www.bioinformatics.vg/bioinformatics_tools/JVT.shtml), which shows all three reading frames of a translated DNA sequence, was particularly useful in monitoring correct protein sequence following a change in reading frame occurring as a result of an intron. Conservations of amino acid sequences at intron/exon boundaries helped to prevent erroneous splice sites from being predicted.

Identified introns were generally of similar length (45 to 86 nucleotides) for StzA orthologues. A worked example showing the methods used to correctly annotate *A. fumigatus* intron 2 is shown in Figure 5.2.

On occasions, computer annotation resulted in the incorrect identification of one of the two intron splice sites (5'-GU… or …AG-3'), resulting in introns of incorrect length. For example, an StzA protein alignment revealed that computer annotation resulted in large overestimations of intron 2 for both *Pyrenophora tritici-repentis* and *Neurospora crassa*, where an additional 25 and 21 amino acids were removed from their respective proteins along with the intron. Canonical 5'-GU and AG-3' splice sites could be readily identified in the expected positions of the genomic versions of the *stzA* DNA sequences so that the appropriate amino acids could be restored to the protein sequences. It was also evident that the length of *A. terreus* intron 2 had been underestimated due to the presence of an erroneous 7 amino acids extension in the protein alignment, and therefore it was necessary to remove the corresponding nucleotides from the start of the intron. The correct 3' splice site was again verified by examination of the nucleotide sequence.

At NCBI, intron 4 was always correctly identified in *stzA* orthologues that possessed this intron. Correct annotation was probably aided by the high degree of homology surrounding this intron as a result of its position within the well-characterised zinc finger domain. Surprisingly, intron 3 in *A. oryzae stzA* was not identified at NCBI despite the fact that this intron was identifed in the *A. flavus stzA* orthologue (N.B. *stzA* sequences between these two orthologues are 100% identical).

In summary, manual annotation of *stzA* genes/StzA proteins revealed that 18 out of 86 introns failed to be identified at NCBI, and a further 3 were misannotated (Intron 2 sequences of *P. tritici-repentis* and *N. crassa* were greatly overestimated in length and intron 2 of *A. terreus* was underestimated in length.). Sixteen of the 27 protein sequences annotated solely by computer analysis (all sequences excluding *A. nidulans* and *T. reesei*) differed in sequence following manual annotation (Tables 5.4 and 5.5).

Table 5.4. A summary of manual annotations to computer annotated StzA protein sequences at NCBI.

Species	Predicted protein size (aa)	Annotated protein size (aa)	mRNA sequence changes	Protein sequence changes
Aspergillus nidulans	698	698	None	None
Aspergillus fumigatus	888	825	Deletion of intron 1: GUAAAGGUAUUUUUUUUUAUCUUUUGAACCUGCAGCCUCAUUUUCUUUUUGCUAACAUGCUUGUUAUAG Deletion of intron 2: GUGAGUGAUUUUCAAAAAAGGAAUGGCGUUGGGUGCAUUGCUCAUUCAUUUUCUCGUGUCUUGAAG Deletion of intron 3: GUAUGUUGUCUCCCCGCCCUCCCGGUGGUUUGCAAGGUUUAACAUUGUCCGCAG	Deletion of intron 1: KGIFFYLLNLQPHFLFANMLVIV (S changed to M) Deletion of intron 2: VSDFQKRNGVGCIAHSFSRVLK Deletion of intron 3: KYVVSPPSRWFARFNIVR (S changed to N)
Aspergillus flavus	724	724	None	None
Aspergillus oryzae	741	724	Deletion of intron 3: GUAUGUUCAUGUCCUUGGUUUCUGGAAGAUGAUGGCUGAUUUCAAUAAUAG	Deletion of intron 3: YVHVLGFWKMMADFNNS
Aspergillus clavatus	817	817	None	None
Aspergillus niger	739	739	None	None
Aspergillus terreus	692	685	Deletion of 21 nucleotides from the start of intron 2: GUACGAAGCGUUACCCAGGUU	Deletion of 7 amino acids from the start of intron 2: VRSVTQV
Neosartorya fischeri	845	825	Deletion of intron 3: GUAUGUUCUCUUCCCGCCCUCUCGGUGGUUUGCAAGGUUUAACAUUGUCCGCAG	Deletion of intron 3: YVLFPPSRWFARFNIVRS (K changed to L)
Talaromyces emersonii	746	731	Deletion of intron 3: GUCUUGUCUGGCCGCUGAUAAUCGUGGUGUACUGACGCGUCCUAG	Deletion of intron 3: SCLAADNRGVLTRPS
Penicillium chrysogenum	824	824	None	None
Arthroderma benhamiae	849	828	Deletion of intron 1: GUUCGUUGCAUAACAUAACCCGGAAUGUUGCAUCUUCUUUGCUGAAUACUAACUAUCUUGCAG	Deletion of intron 1: GSLHNITRNVASSLLNTNYLA
Arthroderma gypseum	836	836	None	None
Trichophyton rubrum	828	828	None	None
Paracoccidioides brasiliensis	889	869	Deletion of intron 3: GUAUGUGCAAAAAUCCCUAGGAAAUAAUACCCUUUCUGCCUUGUGCUCACAUGUUUUCAG	Deletion of intron 3: YVQKSLGNNTLSALCSHVFS
Uncinocarpus reesii	656	639	Deletion of intron 1: GUGCAACACCCGCUUCUUUCAUCCCAUCGCUCAGGCUCAAUGACCAUGGAG	Deletion of intron 1: VQHPLLSSHRSGSMTME
Coccidioides immitis	789	789	None	None
Ajellomyces capsulatus	919	897	Deletion of intron 2: GUGAGUCAGAUUCGCAACCCUCAUAAUGUACCGGACGUUGCUUAUACCAUCAAUCUUCAUUUAAAG	Deletion of intron 2: VSQIRNPHNVPDVAYTINLHLK
Hypocrea jecorina	733	733	None	None
Gibberella zeae	741	741	None	None
Magnaporthe grisea	737	737	None	None
Verticillium albo-atrum	583	563	Deletion of intron 2: GUAAGUCAACUCUUGUAUCAACGCAGCAUUUGGGGAGAUACUUAUCGGUCGUUGAUGAAG	Deletion of intron 2: VSQLLYQRSIWGDTYRSLMK
Nectria haematococca	742	742	None	None
Sordaria macrospora	774	735	Deletion of intron 2: GUAAGCGAAACUGUUGAUCCACUGGACGUUAGGAUGACUACUUACCGUGCGCUACUUUUAAAG Deletion of intron 3: GUACGACUCUCUCUUCUCUUGCAACUGUGUGGCUUGCGCUGACAUUUACUUUAG	Deletion of intron 2: VSETVDPLDVRMTTYRALLLK Deletion of intron 3: YDSLFSCNCVACADIYFS
Podospora anserina	740	740	None	None
Chaetomium globosum	681	663	Deletion of intron 3: GUAUGGCUCUCUUAUCCCUUCUGAUCUAGCCCGCCUUGCUGACUUAGCCUCCAG	Deletion of intron 3: YGSLIPSDLARLADLASS
Phaeosphaeria nodorum	767	750	Deletion of intron 1: GUCCAUCGCAGCUACCCGUCUUGGGAGCCCCGCCUUCUGACAUACUUAAAG	Deletion of intron 1: RPSQLPVLGAPPSDILK (D changed to H)
Botryotinia fuckeliana	722	734	Addition of 96 nucleotides at the gene start: AUGUCGUUCACACACCCCCGAAGAACUACAAAGACCACAUCGAUUGCUGUGACUCUUCUUGCGAAUUCGUCAAGCCGGCCAACAGUUAAUACAACU Deletion of intron 2: GUAAGUGAAAUCCAUAAUGAUGUUGCUUGGGAGUUUACUCACUGGUUCUCGCAAUUGAAG	Addition of 32 amino acids at the protein start: MSFTHPRRTTKTTSIAVTLLANSSSRPTVNTT Deletion of intron 2: VSEIHNDVAWEFTHWFSQLK
Neurospora crassa	714	735	Addition of 63 nucleotides after intron 2: GAGAGGACCAAAGCUGCAGGCUCAUACCUCGAUUUCUGUCUAACGACAAUCGACUGCCUCCAG	Addition of 21 amino acids after intron 2: ERTKAAGSYLDFCLTTIDCLQ
Pyrenophora tritici-repentis	750	736	Deletion of intron 1: GUAAGCACAUUAUGCAGCCAUCCGGCUUAGAUCAGGGCGCCCCGCAUUCUAACAUAUAUAAAG Addition of 75 nucleotides after intron 2: GAGCUGUCGGCUACAUCGGACUCGUACUCCCGCUUCUGCCAACGAUCAAUUCAACUAUUGGCUACGACCGUUCAG Deletion of intron 3: GUAUGACCUCGCCGAUGGCUACUUGUUUGGACGCACUAACAGACUGUUCGGUAG	Deletion of intron 1: GKHIMQPSGLDQGAPHSNIYK Addition of 25 amino acids after intron 2: ELSATSDSYSRFCQRSIQLLATTVQ Deletion of intron 3 YDLADGYLFGRTNRLFGS

Figure 5.2. Manual annotation of *A. fumigatus* StzA intron 2.

Computer predictions at NCBI and the Broad Institute revealed that *Aspergillus* StzA proteins had an intron conserved in position (intron 2) in their N-termini (except StzA from *A. nidulans,* which has been experimentally characterised and is known to lack intron 2). Hence, it seemed surprising that computer annotation did not predict an intron within this region for *A. fumigatus* at NCBI or the Broad Institute. This worked example shows the steps that were taken to manually determine if the *A. fumigatus* gene was likely to possess intron 2 and, if so, to deduce its sequence. The procedure involved the use of (a) an *Aspergillus* StzA protein alignment, (b) an *stzA* genomic DNA alignment and (c) knowledge of fungal intron structure.

5.2a. Partial *Aspergillus* StzA protein alignment (ClustalW)

```
                                        Intron 2
                                       |----------------------|
A.fumigatus      QRIVDREIGCLRDLEKTLLWLAPVSDFQKRNGVGCIAHSFSRVLKNYAASRASYLRFCEFTIQCLHTSVYHL 352
N.fischeri       QRIVDREIGCLRDLEKTLLWLAP----------------------NYAASRASYLRFCEFTIQCLHTSVYHL 307
A.clavatus       QRIVSKEINCLRDLEKTLLWLAP----------------------HYATTRASYIGFCEFTIQCLYSTVQHL 300
A.oryzae         QQIEKQQLKCLRDVEKTLLFSTP----------------------DVKAPTAAWYSFCRFTIHCLHETSGYL 277
A.flavus         QQIEKQQLKCLRDVEKTLLFSTP----------------------DVKAPTAAWYSFCRFTIHCLHETSGYL 277
A.nidulans       QQIDDERISCLRDLEKTVFSLAP----------------------EVKTNDAAYVRFCQYTILCLGQTVSFL 293
A.niger          EQIEKKQIGTLRDLEKSLLHGAA----------------------DVEAEVGSYYQFCSSTILCLSETYTHL 284
                 ::* ...:  ***:**:::  :.                      .  :  .::  **  ** **  :  .*
```

The *Aspergillus* StzA protein alignment (Figure 5.2a) shows an amino acid extension for *A. fumigatus* (highlighted in green) within the well conserved region surrounding intron 2, revealing that this is likely to represent an unrecognised intron in this species.

5.2b. Partial *Aspergillus* StzA genomic DNA alignment (ClustalW)

```
A.fumigatus      ACCTCGAAAAGACACTCTTGTGGCTGGCACCGGTGAGTGATTTTCAAAAAAGGAATGGCG 937
N.fischeri       ACCTCGAAAAGACACTCTTGTGGCTGGCACCGGTAAGTGATTTTCAAAAAAGGAATGGCG 931
A.clavatus       ATCTTGAAAAGACACTCCTGTGGCTCGCTCCGGTAAGTGGATTTCAGGCAA---AAGGCG 909
A.oryzae         ATGTTGAAAAGACCCTTCTCTTTTCCACTCCTGTAAGT---TGCCCGCCATAGTAACCCG 830
A.flavus         ATGTTGAAAAGACCCTTCTCTTTTCCACTCCTGTAAGT---TGCCCGCCATAGTAACCCG 830
A.niger          ATCTCGAGAAGTCTCTTCTCCACGGAGCTGCTGTAAGTACAATTTTAACGTGAGCTTAAA 856
                 *  * **.***:* **  *       .*: * **.***   :    . ..:   .:  ..

A.fumigatus      TTGGGTGCATTGCTCATTCATTTTCTCGTGTCTTGAAGAACTACGCGGCTTCCAGAGCTT 997
N.fischeri       TTAGGTGCATTGCTCATTTATTTTCTCGTTTCTTGAAGAACTACGCGGCTTCCAGAGCTT 991
A.clavatus       TTTGGTGCACTGCTCATTGTTTTTATT-CGGTTTGAAGCATTATGCGACCACCAGGGCGT 968
A.oryzae         TTGAG---------CATCGCTTATATCGTTTCTAGGACGTGAAAGCGCCCACTGCAGCCT 881
A.flavus         TTGAG---------CATCGCTTATATCGTTTCTAGGACGTGAAAGCGCCCACTGCAGCCT 881
A.niger          TTTGTG---------TGCGACTAACAAAACTGCAGGACGTGGAAGCGGAGGTTGG---CT 904
                 ** .          :     *::.:       :*.*  :  * *** .    .     *
```

The genomic DNA sequence incorporating intron 2 for *Aspergillus* species shows introns (highlighted) predicted by computer analysis (Figure 5.2b). The sequence highlighted in green represents the DNA sequence of the amino acid extension in *A. fumigatus*, which is likely to represent a true intron since its length and sequence shows conservation with those from other *Aspergillus* species (highlighted in yellow). Since *A. fumigatus* StzA is 95% similar to *N. fischeri* StzA (Table 5.8 in Section 5.2.4), it is not surprising that the alignment shows that a comparison of intron 2 from these two species reveals identical sizes (66 bp) and almost identical sequences (62/66 bases conserved). Since 66 bp is divisible by 3, the correct reading frames are maintained following the removal of these introns.

5.2c. Intron structure

```
A.fumigatus      GUGAGUGAUUUUCAAAAAAGGAAUGGCGUUGGGUGCAUUGCUCAUUCAUUUUCUCGUGUCUUGAAG
N.fischeri       GUAAGUGAUUUUCAAAAAAGGAAUGGCGUUAGGUGCAUUGCUCAUUUAUUUUCUCGUUUCUUGAAG
A.oryzae           GUAAGUUGCCCGCCAUAGUAACCCGUUGAGCAUCGCUUAUAUCGUUUCUAG
A.flavus           GUAAGUUGCCCGCCAUAGUAACCCGUUGAGCAUCGCUUAUAUCGUUUCUAG
A.niger          GUAAGUACAAUUUUAACGUGAGCUUAAAUUUGUGUGCGACUAACAAAACUGCAG
A.terreus          GUACGAAGCGUUACCCAGGUUGUAUGGUGCAUGGCUAACUUGUAGCAG
A.clavatus       GUAAGUGGAUUUCAGGCAAAAGGCGUUUGGUGCACUGCUCAUUGUUUUUAUUCGGUUUGAAG
```

Manually predicted *A. fumigatus* intron 2, like other intron 2 sequences predicted by computer analysis, contains putative 5' and 3' fungal consensus splice sites that closely match those identified for fungi (GURWGU and YAG, respectively; Kupfer *et al*., 2004). These splice sites are underlined in Figure 5.2c, where nucleotides matching the consensus are highlighted in red. The fungal branchpoint consensus sequence has been identified as RCURAY (Kupfer *et al*., 2004). The *A. fumigatus* example, like other introns, possesses a very similar branchpoint sequence near the 3' end of the intron (shown in the box, where sequences are aligned by the fungal branchpoint: **A**). The position of the branchpoint sequence is typical considering that the mean distance of the fungal branchpoint has been shown to be located between 75–87% of the total intron length (Rep *et al*., 2006). R = adenine or guanine; W = adenine, thymine or uridine; Y = thymine, uridine or cytosine.

Table 5.5. Intron positions in *stzA* orthologues. Protein sizes (shown in aa) resulting from computer annotation and following manual annotation are indicated by numbers to the left and right of the arrow, respectively. Introns 1–4 refer to the known conserved positions of introns within *stzA* orthologues. Dark green boxes indicate protein sequences at NCBI that changed as a result of manual annotation. Note that the manually annotated *B. fuckeliana* protein (*) appears larger following the deletion of intron 2. This is because the N-terminus (32 aa) of this protein was incorrectly omitted as a result of computer annotation and was replaced during the manual annotation process. Intron positions and sizes (shown in bp) are manually corrected versions shown within genomic versions of the *stzA* gene. Pale green boxes indicate introns that were identified or changed as a result of manual annotation and therefore differ from the original annotations ([a] Intron not identified by computer annotation. [b] Intron size underestimated by computer annotation. [c] Intron size overestimated by computer annotation). StzA proteins annotated at the Broad Institute have additional missed introns or incorrectly annotated introns compared to those annotated at NCBI, and such introns are indicated by ([d]) and highlighted by pale red boxes. Dark red boxes highlight proteins incorrectly annotated at the Broad Institute but correctly annotated at NCBI. An additional intron appeared to be wrongly predicted in the C-terminus of *G. zeae* StzA at the Broad Institute but is not shown. The *T. reesei*, *T. emersonii*, *P. chrysogenum*, *P. anserina*, *N. haematococca*, *S. macrospora* and *V albo-atrum* StzA orthologues were not available at the Broad Institute. Dark blue boxes indicate the StzA proteins from *A. nidulans* (O'Neil *et al*., 2002) and *T. reesei* (Aro *et al*., 2003) for which introns sequences (light blue) have been experimentally validated.

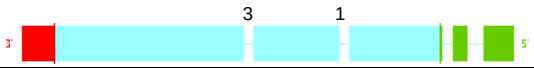
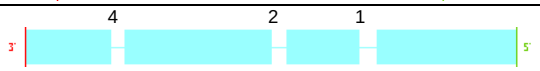
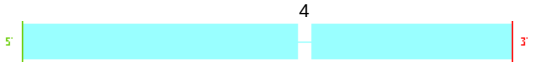
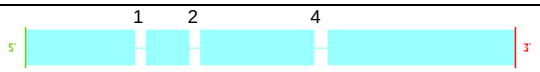
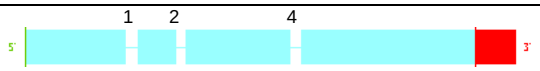
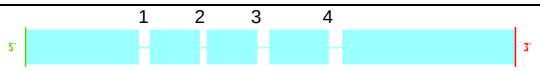
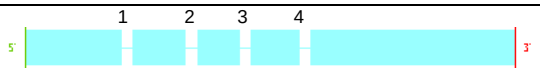
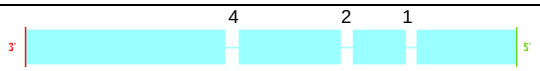
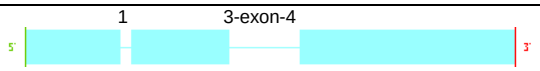
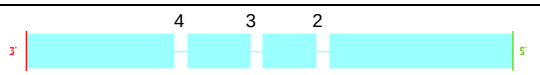
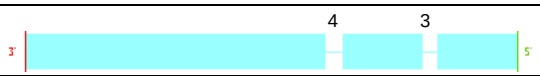
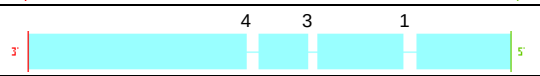
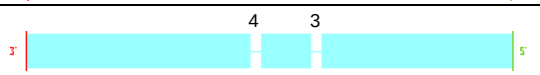
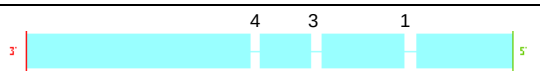
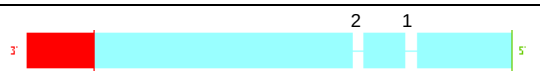
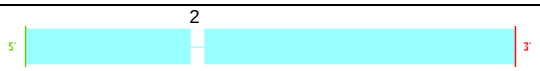
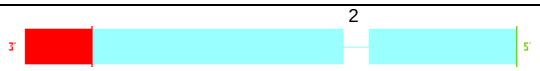
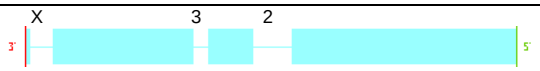
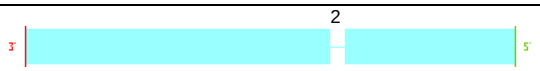
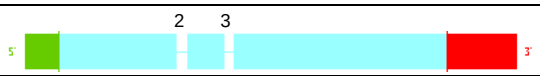
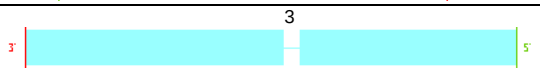
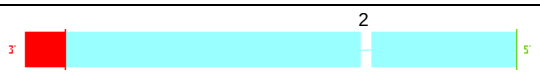
Species	Protein size (aa) Predicted → Annotated	Intron position (size in bp)			
		1	2	3	4
Aspergillus nidulans	698	521–571 (57)		1068–1121 (54)	
Aspergillus niger	739 → 722	506–562 (57)	829–882 (54)	1092–1142 (51) [d]	1416–1464 (49)
Aspergillus fumigatus	888 → 825	548–616 (69) [a]	910–975 (66) [a]	1224–1277 (54) [a]	1548–1620 (73)
Aspergillus oryzae	741 → 724	539–593 (55)	806–856 (51)	1093–1143 (51) [a]	1417–1476 (60)
Aspergillus flavus	741 → 724	539–593 (55)	806–856 (51)	1093–1143 (51) [d]	1417–1476 (60)
Aspergillus terreus	692 → 685	530–578 (49)	797–844 (48) [b]	1078–1133 (56)	1410–1470 (61)
Aspergillus clavatus	817	533–591 (59)	885–946 (62)	1183–1234 (52)	1508–1566 (59)
Neosartorya fischeri	845 → 825	554–610 (57)	904–969 (66)	1218–1271 (54) [a]	1542–1615 (74)
Ajellomyces capsulatus	919 → 897	581–648 (68)	927–992 (66) [a]	1241–1306 (66) [d]	1592–1670 (79) [d]
Uncinocarpus reesii	656 → 639	562–612 (51) [a]	799–854 (56)	1094–1144 (51)	1424–1479 (56)
Talaromyces emersonii	746 → 731	545–595 (51)	865–927 (63)	1167–1211 (45) [a]	1485–1549 (65)
Coccidioides immitis	530 → 789	563–648 (86) [d]	906–965 (60) [d]	1196–1246 (51)	1526–1582 (57)
Penicillium chrysogenum	824	512–591 (80)	846–924 (79)	1170–1222 (53)	1499–1557 (59)
Arthroderma benhamiae	849 → 828	524–586 (63) [a]		1044–1098 (55)	1375–1431 (57)
Arthroderma gypseum	836	524–595 (72)		1077–1129 (53)	1406–1467 (62)
Trichophyton rubrum	828	524–586 (63)		1044–1098 (55)	1375–1429 (55)
Paracoccidioides brasiliensis	889 → 869	632–704 (73)	995–1057 (63)	1306–1365 (60) [a]	
Phaeosphaeria nodorum	767 → 750	482–532 (51) [a]	823–888 (66)	1125–1178 (54) [d]	
Pyrenophora tritici-repentis	750 → 736	479–541 (63) [a]	829–897 (69) [c]	1134–1187 (54) [a]	
Trichoderma reesei	733		736–798 (63)	1026–1091 (66)	
Gibberella zeae	461 → 741		748–810 (63) [d]	1026–1076 (51)	
Chaetomium globosum	681 → 663		733–795 (63)	1014–1067 (54) [a]	
Neurospora crassa	714 → 735		706–768 (63) [c]	996–1049 (54)	
Podospora anserina	740		733–795 (63)	1020–1076 (57)	
Botryotinia fuckeliana	722 → 734*		784–843 (60) [a]	1080–1148 (69)	
Nectria haematococca	742		751–813 (63)	1029–1079 (51)	
Sordaria macrospora	774 → 735		706–768 (63) [a]	996–1049 (54) [a]	
Verticillium albo-atrum	583 → 563		745–804 (60) [a]		
Magnaporte grisea	737		727–783 (57)		

StzA orthologues could also be found at the Broad Institute (http://www.broad.mit.edu/node/568) for species sequenced as part of the Fungal Genome Initiative (Cuomo and Birren, 2010). The vast majority of StzA orthologues could be located at the Broad Institute, with the exception of *T. reesei*, *Talaromyces emersonii*, *Penicillium chrysogenum*, *Podospora anserina*, *Nectria haematococca*, *Sordaria macrospora* and *Verticillium albo-atrum*. Sequence alignments were performed to ensure that the genomic version of each *stzA* gene was identical to the corresponding sequence on the NCBI website. These sequences were identical but the species *Gibberella zeae*, *Phaeosphaeria nodorum*, *Botryotinia fuckeliana*, *Ajellomyces capsulatus* and *Microsporum gypseum* were referred to by the names of their corresponding anamorphs/telomorphs (i.e. *Fusarium graminearum*, *Stagonospora nodorum*, *Botrytis cinerea*, *Histoplasma capsulatum* and *Arthroderma gypseum*, respectively). *Aspergillus niger stzA* had its 3' end missing (600 bp), which was noted by the Broad Institute website.

All *stzA* orthologues at the Broad Institute (October 2012) possessed the same intron annotation mistakes as those identified at NCBI (with the sole exception of *N. crassa* intron 2 being wrong at NCBI but correct at the Broad Institute), but additional mistakes were also evident, indicative of differences in annotation methods between these two websites (Figure 5.6). An additional 8 introns were misannotated at the Broad Institute, leading to an additional 4 incorrect protein sequence predictions. In *G. zeae*, intron 2 was overestimated in size by 189 bp, resulting in the deletion of 63 amino acids from the predicted translated protein. Furthermore, an erroneous intron appeared to be incorrectly predicted in the C-terminus of *G. zeae*, immediately following the zinc finger region. This would introduce a STOP codon at the beginning of the C-terminus, and although unidentified introns might be present in the C-termini of StzA proteins, the occurrence of this intron would substantially truncate this protein (from 741 to 461 aa in length), so was considered to be incorrect considering that the shortest StzA protein (*V. albo-atrum*) is 563 amino acids in length. For *A. capsulatus*, introns 3 and 4 and the intervening exon were treated as one long intron (430 bp) and removed from the protein sequence, resulting in the loss of 73 amino acids from the predicted translated protein. The StzA protein alignment proved useful in identifying the omission of the start of the *B. fuckeliana* StzA protein in addition to its usefulness in intron identification. The corresponding genomic DNA was used to retrieve the 32 amino acids that were omitted. For the *Coccidioides immitis* StzA protein, 923 bases were missing from the N-terminus, truncating the StzA protein from 789 to 530 amino acids, preventing introns 1 and 2 from being predicted by computer analysis. A summary of the misannotated gene/protein StzA sequences at the Broad Institute are shown in Table 5.6, which reveals that out of a total of 86 introns, 21 and 28 introns were

found to be misannotated at NCBI and the Broad Institute, respectively. Following manual annotation, 16 (out of 29) and 19 (out of 22) *stzA* genes and their corresponding proteins differed in sequence at NCBI and the Broad Institute, respectively (Tables 5.5 and 5.6).

Table 5.6. Intron misannotations in StzA orthologues at the Broad Institute. Solid pale blue segments represent exons and breaks represent predicted introns. Numbers represent conserved intron positions (1–4). Green and red segments represent sequenced regions outside the gene, in the 5' and 3' terminal regions, respectively. Many introns within *stzA* genes were annotated incorrectly. Furthermore, in addition to intron annotation errors, other mistakes were evident. The N-termini were missing from the *stzA* genes of *C. immitis* (923 bp) and *B. fuckeliana* (96 bp). The final 200 amino acids from the C-terminus of *A. niger* were not available for analysis. An erroneous intron was predicted in the C-terminus of *G. zeae* StzA. This intron truncated the protein from 741 to 461 amino acids by introducing a premature STOP codon, indicated by (X).

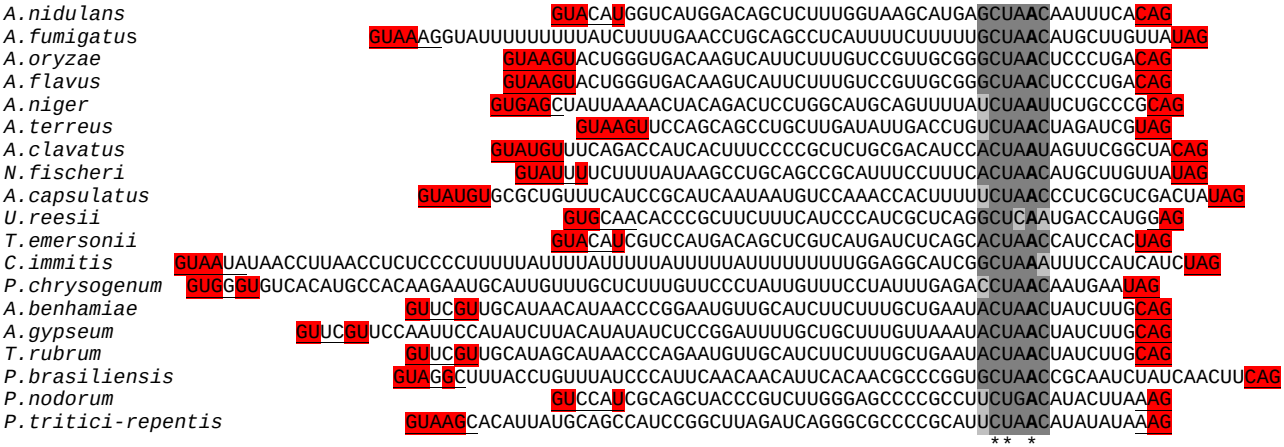
Species	Accession number	Protein size (aa) Predicted → Annotated	Broad Institute intron predictions for *stzA* genes	Assessment of intron predictions
Aspergillus nidulans	ANID_02919.1	698		Both introns correctly identified.
Aspergillus niger	est_GWPlus_C_70812	722 → 739		Intron 3 not identified. Missing 3' end (600 bp).
Aspergillus fumigatus	Afu3g08010	888 → 825		Introns 1, 2 and 3 not identified.
Aspergillus oryzae	AO090005001502	741 → 724		Intron 3 not identified.
Aspergillus flavus	AFL2G_01400.2	741 → 724		Intron 3 not identified. Introns 1 and 4 incorrectly identified.
Aspergillus terreus	ATEG_01689.1	692 → 685		Intron 2 underestimated in size by 21 bp.
Aspergillus clavatus	ACLA_035940	817		All 4 introns correct.
Neosartorya fischeri	NFIA_069130	845 → 825		Intron 3 not identified.
Ajellomyces capsulatus (*Histoplasma capsulatum*)	HCBG_06660.2	824 → 897		Intron 2 not identified. Exon removed between introns 3 and 4.
Uncinocarpus reesii	UREG_01403.1	656 → 639		Intron 1 not identified.
Coccidioides immitis	CISG_02567.1	530 → 789		N-terminus (923 bases) missing. Introns 1 and 2 not predicted.
Arthroderma gypseum (*Microsporum gypseum*)	MGYG_05167.2	836		All 3 introns correct.
Arthroderma benhamiae	ARB_06580	849 → 828		Intron 1 not identified.
Trichophyton rubrum	TERG_02061.2	828		All 3 introns correct.
Paracoccidioides brasiliensis	PAAG_01715.1	889 → 869		Intron 3 not identified.
Phaeosphaeria nodorum (*Stagonospora nodorum*)	SNOG_15249.1	785 → 750		Introns 1 and 3 not identified.
Pyrenophora tritici-repentis	PTRG_07931.1	750 → 736		Introns 1 and 3 not identified. Intron 2 overestimated in size by 75 bp.
Gibberella zeae (*Fusarium graminearum*)	FGSG_11792.3	461 → 741		Intron 2 overestimated in size by 63 bp. Intron predicted (77 bp) following Zn finger region.
Chaetomium globosum	CHGG_00776.1	681 → 663		Intron 3 not identified.
Neurospora crassa	NCU09333	735		Both introns correct.
Botryotinia fuckeliana (*Botrytis cinerea*)	BC1G_11939.1	722 → 734		Intron 2 not identified. N-terminus (96 bp) missing.
Magnaporthe grisea	MGG_04428.6	737		Intron correctly identified.

5.2.3 Intron characteristics

Identified introns were generally of similar length (45 to 86 nucleotides) for StzA orthologues. All 86 introns (Figure 5.3) of the *stzA* orthologues possessed the eukaryotic 5' splice site and 3' splice site consensus sequences GU and AG, respectively (Kupfer *et al*., 2004). There were no GC-AG introns, which have been identified in approximately 1% of the introns of filamentous ascomycetes (Kupfer *et al*., 2004; Rep *et al*., 2006). Fifty-eight out of 86 introns exactly matched the fungal 5' splice site consensus sequence GURWGU and 59 matched the fungal 3' splice site consensus sequence YAG (Table 5.7; Kupfer *et al*., 2004).

Correct annotation of intron sequences was verified by identifying branchpoint sequences. Approximately half of the introns (42 / 86) contained a match for the fungal branchpoint consensus sequence RCURAY (Kupfer *et al*., 2004) and 40 out of the remaining 44 introns deviated from this consensus sequence by just a single nucleotide (Table 5.7). The uracil (U) residue was 100% conserved and the neighbouring cysteine (C) residue was conserved in 79 of the branchpoint sequences and was replaced by a uracil residue in the remaining 7 sequences. In a study of fungal introns by Kupfer *et al*. (2004), the mean distance of the branchpoint lay between 75 and 87% of the total intron length (i.e. the branchpoint lay between nucleotides 75 and 87 for an intron length of 100 nucleotides). Figure 5.3 clearly indicates that the predicted branchpoint sequences for introns in *stzA* orthologues lie within a similar region. Deviations from intron consensus sequences at each intron position are summarised in Figure 5.3.

Intron 1



Intron 2

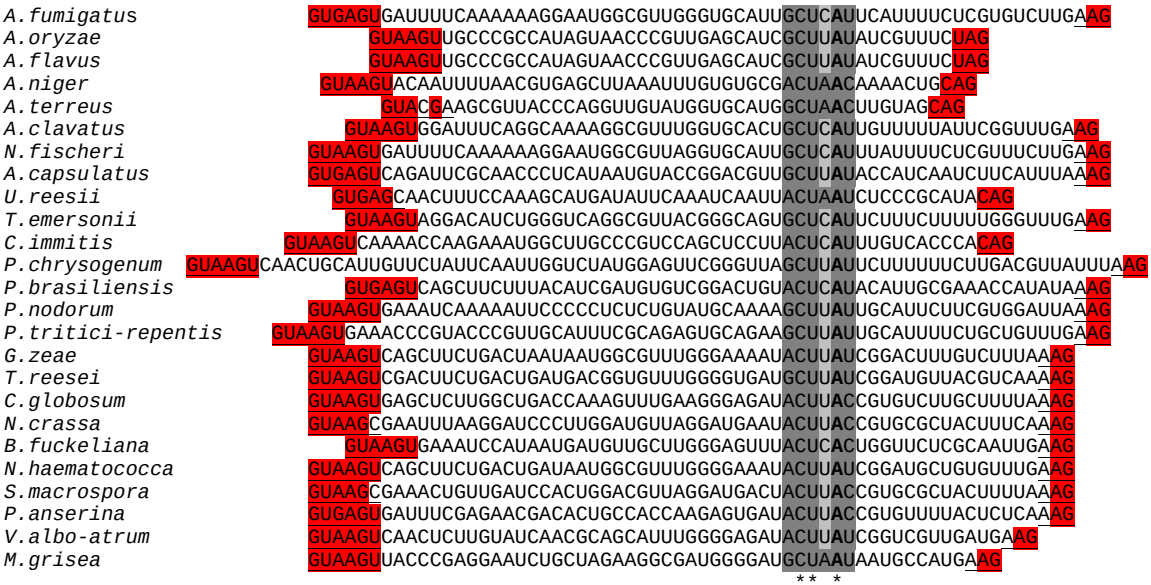
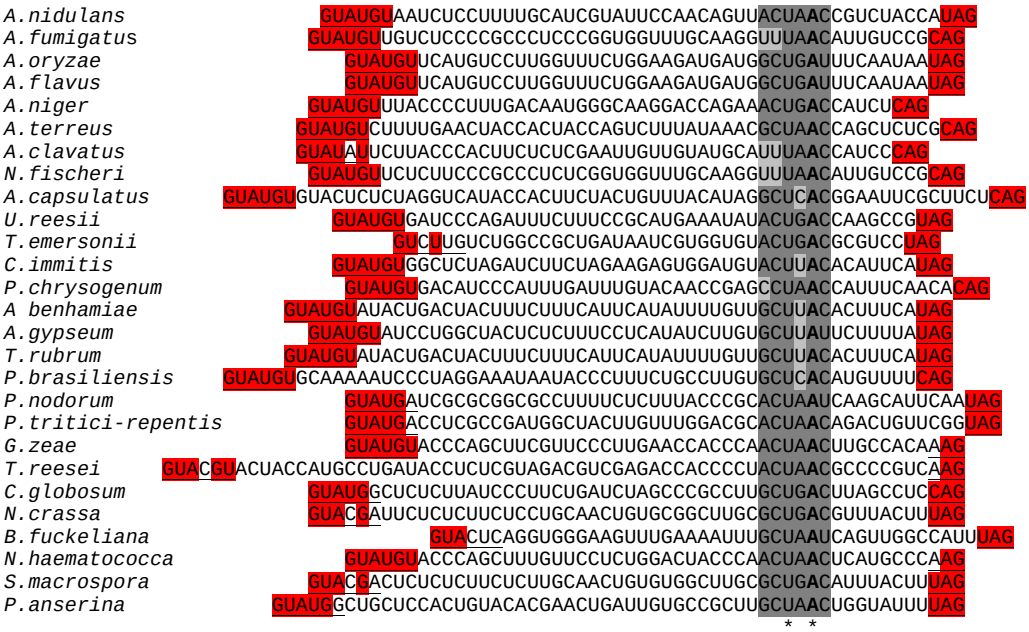


Figure 5.3. Intron sequences of *stzA* orthologues (continued overleaf). The 5' splice sites (6 residues) and 3' splice sites (3 residues) are underlined. Conservation of residues when compared to the 5' and 3' fungal consensus splice sites GURWGU and YAG is indicated by red highlights (R = adenine or guanine, W = adenine, thymine or uridine, Y = thymine, uridine or cytosine). The positions of manually predicted fungal branchpoint sequences are indicated by grey highlights. Residues that conform to the branchpoint sequence RCURAY are shown in dark grey and residues that do not conform to this sequence are shown in light grey. All sequences are aligned by the fungal branchpoint (**A**), shown in bold.

Intron 3



Intron 4

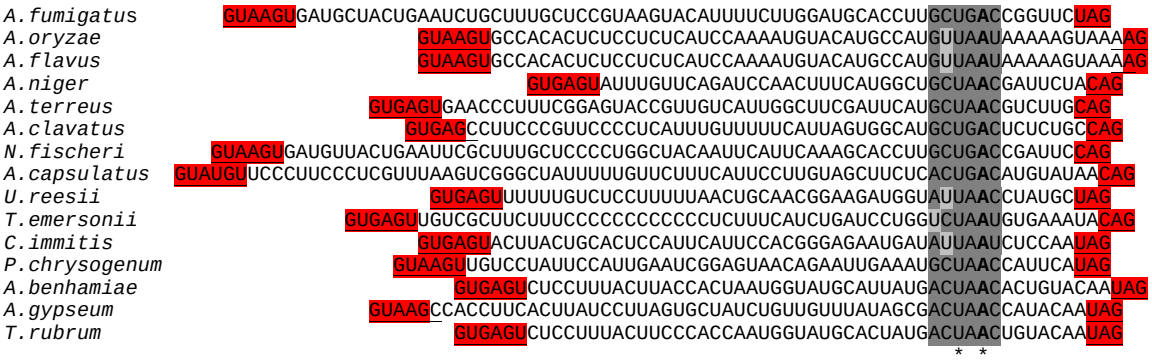


Figure 5.3 (continued).

Table 5.7. Deviations from fungal intron consensus sequences at each conserved intron position. Introns within *stzA* genes are present in four conserved positions and total 86: 19, 25, 27 and 15 introns in positions 1, 2, 3 and 4, respectively. The table shows the extent to which introns contain consensus sequences identified for fungal introns (Kupfer *et al*., 2004). The intron sequences identified are the 5' splice site (GURWGU), the branchpoint (RCURAY) and the 3' splice site (YAG).

Type of intron consensus sequence and extent to which deviation occurs	**Number of sequences which perfectly match the consensus sequence at each intron position**				**Total**
	1	**2**	**3**	**4**	
5' splice site GURWGU (exact)	5 / 19	23 / 25	17 / 27	13 / 15	**58 / 86**
Minus 1 correct nt	4 / 19	1 / 25	6 / 27	2 / 15	**13 / 86**
Minus 2 correct nt	8 / 19	1 / 25	2 / 27	0 / 15	**11 / 86**
Minus 3 correct nt	2 / 19	0 / 25	2 / 27	0 / 15	**4 / 86**
Branchpoint 5'-RCURAY (exact)	11 / 19	4 / 25	17 / 27	10 / 15	**42 / 86**
Minus 1 correct nt	7 / 19	21 / 25	7 / 27	5 / 15	**40 / 86**
Minus 2 correct nt	1 / 19	0 / 25	3 / 27	0 / 15	**4 / 86**
3' splice site YAG (exact)	16 / 19	6 / 25	24 / 27	13 / 15	**59 / 86**
Minus 1 correct nt (NAG)	3 / 19	19 / 25	3 / 27	2 / 15	**27 / 86**

5.2.4 Comparisons of StzA proteins

The final annotated StzA proteins were aligned using Advanced TCoffee (http://tcoffee.vital-it.ch/cgi-bin/Tcoffee/tcoffee_cgi/index.cgi; Appendix 6). The proteins vary in length considerably, from 563 (*V. albo-atrum*) to 897 amino acids (*A. capsulatus*). All proteins are moderately well conserved in the N-terminal region, highly conserved across the zinc finger domain and the region containing a putative nuclear localisation signal (NLS), but poorly conserved in the C-terminus. Thus, most of the protein variability is due to divergence in the C-terminal region, which is highly polymorphic and ranges in length from 69 (*V. albo-atrum*) to 363 amino acids (*A. capsulatus*). For example, *A. nidulans* StzA demonstrates 47% amino acid identity with *A. fumigatus* StzA prior to the zinc fingers, 74% over the zinc fingers, and then only 19% after the zinc fingers.

Percentage identitities between pairs of StzA proteins are shown in Table 5.8. The data clearly show that with regards to sequence, pairs of StzA proteins from organisms of the same fungal class are generally more similar to each other than to proteins from different classes. For example, the *A. nidulans* StzA protein is 42% similar to *A. fumigatus* StzA, 42–32% similar to other *Aspergillus* species and other Eurotiomycetes, but only 28–34% similar to Sordariomycetes. Likewise, comparisons of StzA proteins between any two Sordariomycetes reveal they are always at least 40% identical across their entire proteins. However, StzA proteins within the Sordariomycetes class also demonstrate considerable variability, ranging from 93% similarity for an *N. crassa–S. macrospora* StzA alignment to only 46% similarity for an *M. grisea–V. albo-atrum* StzA alignment. This variability is even more notable for *Aspergillus* StzA comparisons. An *A. nidulans–A. terreus* StzA alignment reveals only 32% similarity, whereas the StzA proteins from *A. fumigatus* and *N. fischeri* are 95% identical.

Table 5.8. Percentage identities between pairs of StzA proteins. Scores were derived from the raw data of a ClustalW alignment of StzA proteins. Comparisons of proteins within classes are shown by coloured highlights: Eurotiomycetes (blue), Sordariomycetes (red) and Dothideomycetes (green).

	A. nidulans	*A. fumigatus*	*N. fischeri*	*A. clavatus*	*A. oryzae*	*A. flavus*	*A. niger*	*A. capsulatus*	*A. terreus*	*T. emersonii*	*C. immitis*	*U. reesii*	*P. chysogenum*	*A. benhamiae*	*A. gypseum*	*T. rubrum*	*P. brasiliensis*	*P. nodorum*	*P. tritici-repentis*	*T. reesei*	*M. grisea*	*C. globosum*	*N. crassa*	*G. zeae*	*P. anserina*	*S. macrospora*	*N. haematococca*	*V. albo-atrum*	*B. fuckeliana*
A. nidulans	-	42	42	41	41	41	39	33	32	39	36	33	36	30	31	30	36	30	30	30	30	30	29	29	28	29	29	34	29
A. fumigatus	42	-	95	70	40	40	38	42	35	54	43	42	50	36	36	36	43	35	36	33	30	35	33	33	32	33	34	39	32
N. fischeri	42	95	-	70	42	42	36	42	36	55	42	43	51	37	37	37	44	36	34	34	31	35	33	34	32	32	33	39	32
A. clavatus	41	70	70	-	43	43	38	42	36	52	46	41	49	35	35	35	42	34	36	34	34	35	29	32	32	30	32	27	33
A. oryzae	41	40	42	43	-	100	38	37	38	42	39	38	38	31	31	31	36	30	31	28	29	32	30	29	30	30	29	34	30
A. flavus	41	40	42	43	100	-	38	37	38	42	39	38	38	31	31	31	36	30	31	28	29	32	30	29	30	30	29	34	30
A. niger	39	38	36	38	38	38	-	30	32	35	30	35	33	27	28	27	33	27	26	32	27	30	27	28	27	27	28	32	27
A. capsulatus	33	42	42	42	37	37	30	-	30	52	53	48	38	43	44	44	63	38	36	33	35	37	34	34	35	33	34	38	35
A. terreus	32	35	36	36	38	38	32	30	-	32	31	33	32	29	28	29	30	27	25	25	24	27	25	27	24	25	27	26	24
T. emersonii	39	54	55	52	42	42	35	52	32	-	45	45	45	39	41	38	50	34	36	32	30	34	31	31	33	32	31	36	33
C. immitis	36	43	42	46	39	39	30	53	31	45	-	63	39	45	47	45	52	35	36	34	32	36	32	33	33	32	35	37	34
U. reesii	33	42	43	41	38	38	35	48	33	45	63	-	37	44	46	44	48	34	35	31	33	34	33	32	35	33	34	32	33
P. chrysogenum	36	50	51	49	38	38	33	38	32	45	39	37	-	34	36	34	41	32	35	30	29	36	31	30	31	32	31	36	29
A. benhamiae	30	36	37	35	31	31	27	43	29	39	45	44	34	-	88	98	42	34	32	30	29	34	30	32	32	31	32	35	28
A. gypseum	31	36	37	35	31	31	28	44	28	41	47	46	36	88	-	88	42	31	33	31	30	34	32	33	32	31	31	35	28
T. rubrum	30	36	37	35	31	31	27	44	29	38	45	44	34	98	88	-	43	33	32	31	29	34	30	32	32	31	32	35	28
P. brasiliensis	36	43	44	42	36	36	33	63	30	50	52	48	41	42	42	43	-	38	37	33	32	35	34	33	35	34	35	37	29
P. nodorum	30	35	36	34	30	30	27	38	27	34	35	34	32	34	31	33	38	-	71	35	31	33	33	29	31	32	32	35	30
P. tritici-repentis	30	36	34	36	31	31	26	36	25	36	36	35	35	32	33	32	37	71	-	32	32	33	33	32	34	32	34	35	32
T. reesei	30	33	34	34	28	28	32	33	25	32	34	31	30	30	31	31	33	35	32	-	48	53	49	65	52	48	70	61	41
M. grisea	30	30	31	34	29	29	27	35	24	30	32	33	29	29	30	29	32	31	32	48	-	56	52	47	55	52	50	46	40
C. globosum	30	35	35	35	32	32	30	37	27	34	36	34	36	34	34	34	35	33	33	53	56	-	64	51	71	63	54	55	47
N. crassa	29	33	33	29	30	30	27	34	25	31	32	33	31	30	32	30	34	33	33	49	52	64	-	47	61	93	51	52	44
G. zeae	29	33	34	32	29	29	28	34	27	31	33	32	30	32	33	32	33	29	32	65	47	51	47	-	52	48	80	59	41
P. anserina	28	32	32	32	30	30	27	35	24	33	33	35	31	32	32	32	35	31	34	52	55	71	61	52	-	60	51	52	43
S. macrospora	29	33	32	30	30	30	27	33	25	32	32	33	22	31	31	31	34	32	32	48	52	63	93	48	60	-	50	53	43
N. haematococca	29	34	33	32	29	29	28	34	27	31	35	34	31	32	31	32	35	32	34	70	50	54	51	80	51	50	-	62	40
V. albo-atrum	34	39	39	37	34	34	32	38	26	36	37	32	36	35	35	35	37	35	35	61	46	55	52	59	52	53	62	-	50
B. fuckeliana	29	32	32	33	30	30	27	35	24	33	34	33	29	28	28	28	29	30	32	41	40	47	44	41	43	43	40	50	-

5.2.5 Sequence conservation among StzA proteins

The zinc finger domain, putative NLS regions and the N- and C-terminal regions of StzA proteins were analysed to assess the extent of sequence conservation within these regions across species.

5.2.5a Zinc finger domain and putative NLS regions

The zinc finger domain, containing three C_2H_2 zinc fingers, is highly conserved across StzA proteins (Figure 5.4). The motif for the zinc finger domain is $CN_{2/4}CN_{12}HN_3HN_5CN_4CN_{15}HN_4HN_8CN_4CN_{12}HN_4H$. Interestingly, organisms that belong to the Eurotiomycetes and Dothidiomycetes classes have two amino acid residues between the two carbon atoms of the first zinc finger, whereas organisms belonging to the Sordariomycetes and the one representative Leotiomycete have four amino acid residues in this region. In *V. albo-atrum*, the second cysteine residue of the third zinc finger is replaced by leucine, indicative of at least one sequencing error in this region.

The most likely region of the bipartite NLS in orthologues of StzA is shown highlighted in green in Figure 5.4. The NLS starts with the two basic amino acids sequence KR and ends with either the four basic amino acids sequence RRKK (26 orthologues) or RRRK (3 orthologues) and is highlighted in green. These sequences are perfectly aligned and are separated by a spacer region of 15–17 amino acids. Furthermore, conservation of individual residues is apparent within this region. The predicted NLS sequences meet the criteria of a bipartite NLS, which is composed of two clusters of basic amino acids composed of lysine (K) and arginine (R): a smaller one of two basic amino acids and a larger one of at least three basic amino acids in a group of five, separated by a spacer of at least ten amino acids (Dingwall and Laskey, 1998).

(a)

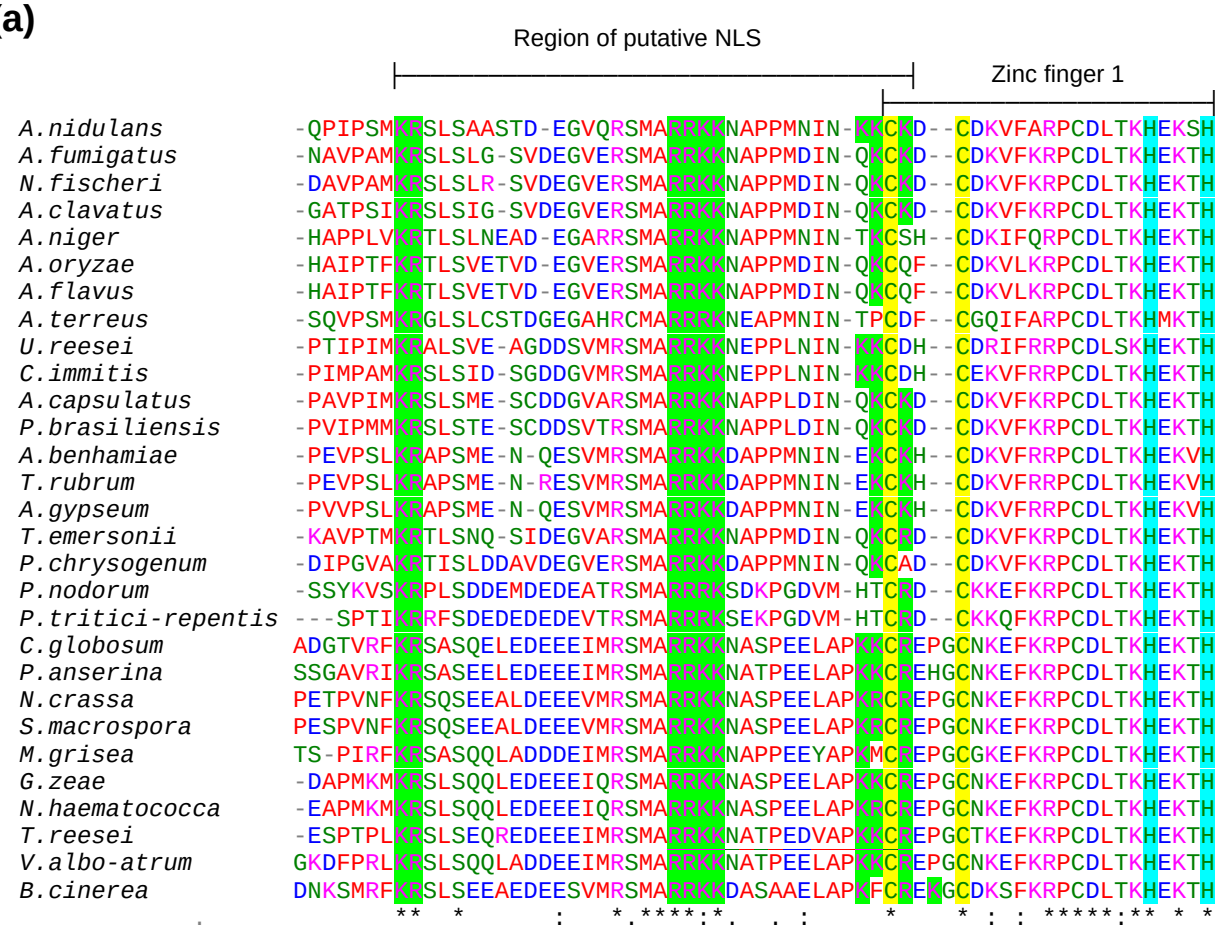
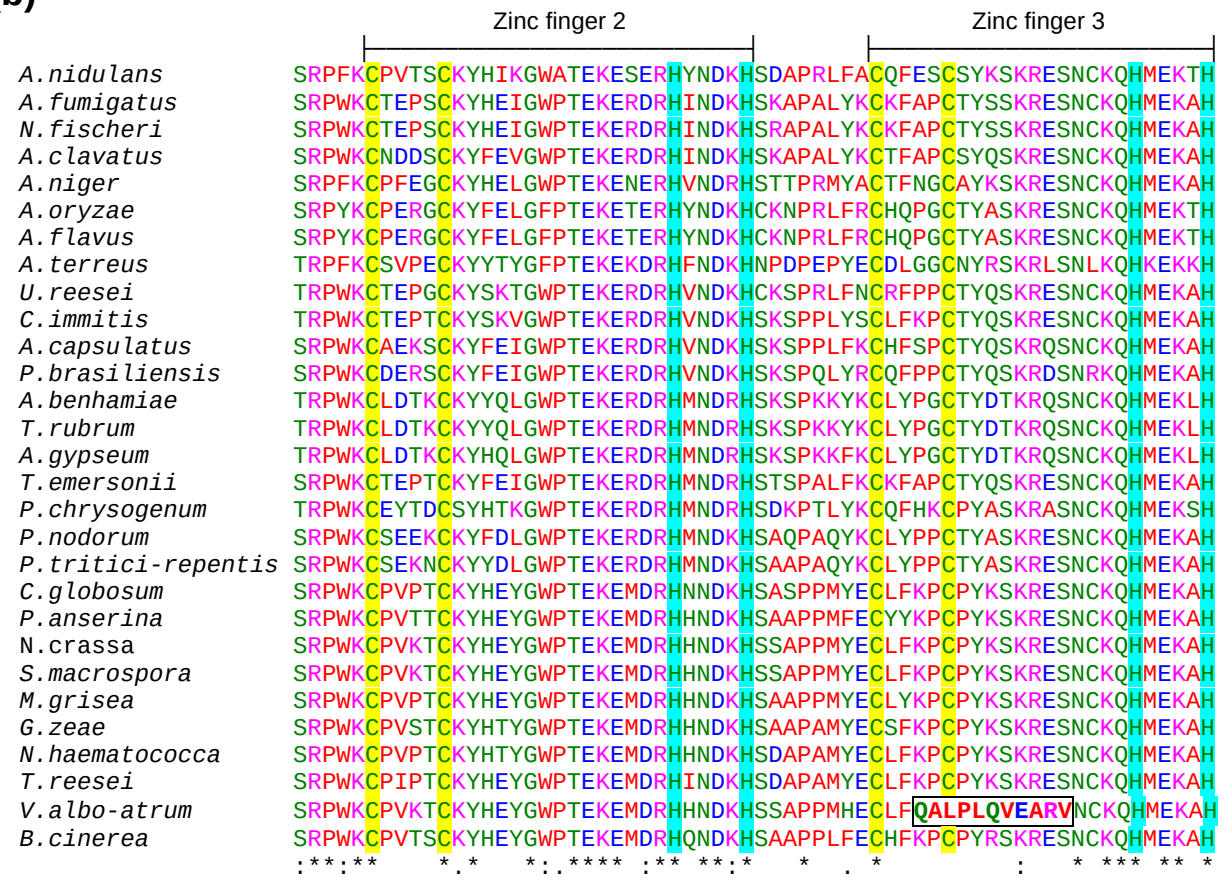


Figure 5.4. ClustalW alignment showing a region containing (a) the putative NLS, (a, b) the zinc finger domain and (c) the start of the C-terminus for all 29 StzA orthologues (continued overleaf). Regions a, b and c form a continuous sequence. Note that the zinc finger region is highly conserved, whereas the C-terminus represents the most divergent region of the proteins. Basic amino acid residues (K and R) within a region likely to contain the bipartite NLS are highlighted in green. The bipartite NLS predicted for *T. reesei* by Saloheimo *et al*. (2000) is underlined (RRKKNATPEDVAPKKCR). The paired cysteine (C) and histidine (H) amino acid residues of the three zinc fingers are highlighted in yellow and blue, respectively. A likely sequencing error in the third zinc finger of the *V. albo-atrum* StzA protein is shown by boxed and highlighted residues. The position of the STOP codon (W) at the start of the C-terminus of the *stzA* gene in the *sltA1* mutant GO281 is indicated by a grey highlight and a black arrow (▼). Red indicates small and hydrophobic residues (including aromatic Y); Blue indicates acidic residues; Pink indicates basic residues; and green indicates hydroxyl, amine and Q. (*) indicates that the residues in that column are identical in all sequences in the alignment. (:) indicates conserved substitutions. (.) indicates semi-conserved substitutions.

(b)



(c)

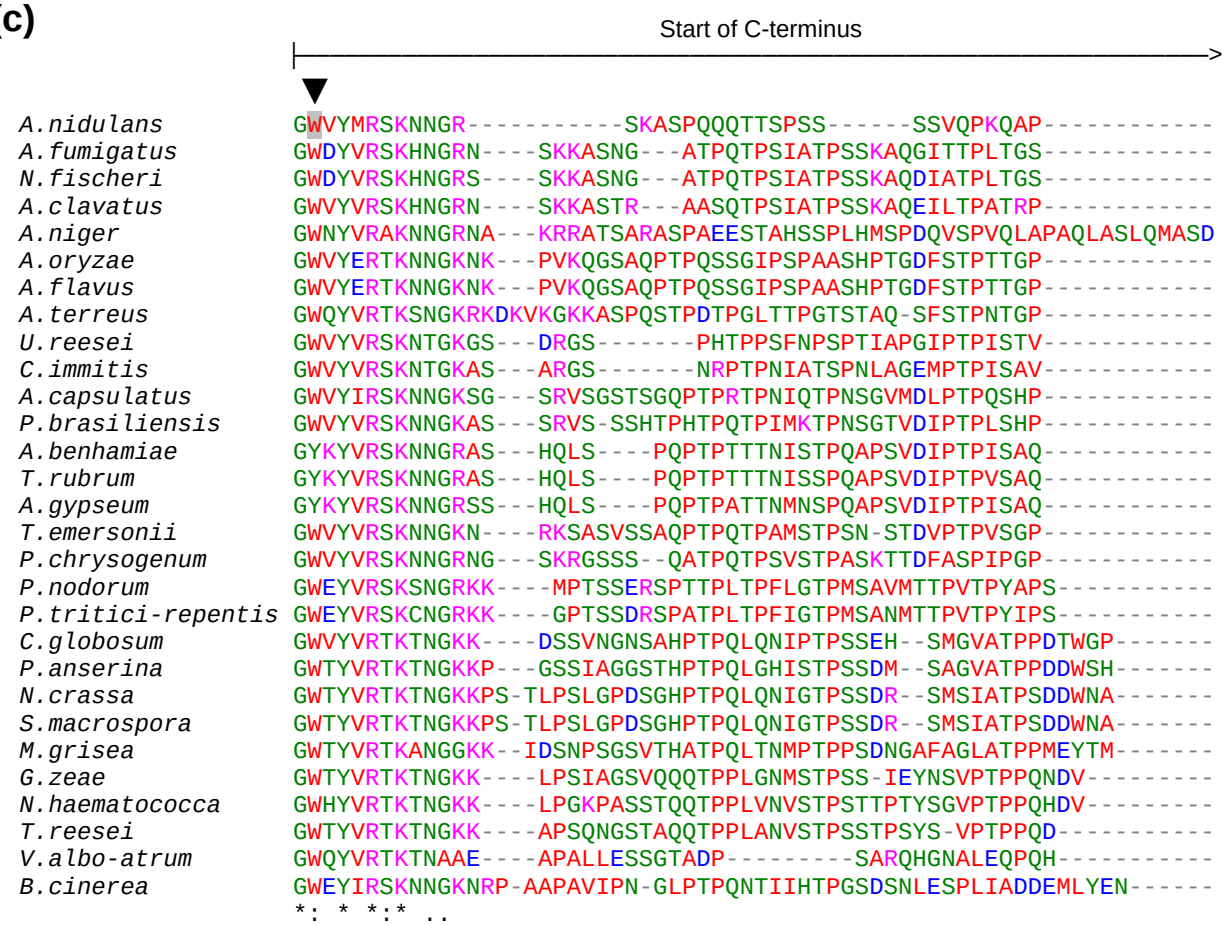


Figure 5.4 (continued).

5.2.5b Conserved sequences in the N-termini

In addition to the conservation evident in the zinc finger domain of StzA proteins and the region of the putative NLS, several short conserved amino acid sequences were apparent within the N-termini of StzA proteins (Table 5.9 and Appendix 6). These sequences, which may have structural and functional significance, included: PRR(R/T), TF(H/S), (S/A)DSG(L/I/V)G(S/T)S(L/I/V), LRD(L/V)E(K/N) and Y(F/I)(L/I/V)DL. Many individual residues may be essential for the maintenance of protein secondary structure as they are 100% conserved across StzA proteins and correlated with predicted conserved alpha helices (Appendix 7).

Table 5.9. Conserved sequences within the N-termini of StzA orthologues. The positions of the conserved sequences are for *A. nidulans* StzA and represent amino acid positions from the start of the protein. Numbers in brackets, shown after sequences, represent the number of species containing that particular sequence, with the maximum number being 29.

Position	Conserved sequence	No. of StzA proteins with this sequence
14–17	PRR(R/T)	27
46–48	TF(H/S)	28
155–163	(S/A)DSG(L/I/V)G(S/T)S(L/I/V)	19
254–259	LRD(L/V)E(K/N)	29
310–314	Y(F/I)(L/I/V)DL	29

In addition to the conserved amino acid sequences in the N-termini of StzA proteins, a raised proportion of alanine residues in this region (close to intron 1) was evident. This was far more pronounced for species belonging to the Sordariomycetes group (Figure 5.5), where the number of alanine residues clearly occurred several times more often than would be expected by chance alone (1 in 20 residues). Alanine rich regions within TFs are known to contribute to transcriptional repression (Tilburn *et al*., 1995), and *T. reesei* Ace1 is a known repressor of cellulase and xylanase expression (Aro *et al*., 2003).

```
C.globosum       ASTNDKDAVAKKDK---A-GR-TTAVTRSAAAR-SATAASAVGLSPRA  191
G.zeae           ASTNEKRGAVTASKEA-K-VQ-TRCLTRSAAAA-AA-TGKLPSLGSKA  197
T.reesei         VSTNDKAGAADSTKK----PQ-ASALTRSAASS---TTAMLPSLSHRA  193
M.grisea         LSTKQKSTPNADSTPAKTGVK-GSAVTRSAAS---STSKNLPSLSARA  190
N.crassa         ASTSEKDASS-KAK---T-TR-TSAVARSATAR-AASTPDLPGLGDRA  183
N.haematococca   ASTNEKHAAVTQSKE----TKTRSAITRSAAAS---SAEKLPSLGSKA  198
P.anserina       ASTNEKIAAK-EQT---V-AK-TTAVTRSAAATRTTTTTTTQVLGQRA  192
S.macrospora     ASTSEKDASC-KAK---T-TR-TSAVTRSAAAR-SASALNLPGLSDRA  183
V.albo-atrum     ASSVEKQAPSITSKT----SK-ASAITRSAAAP-SNTMTKVSGLSSKA  196
                  *: :*       .      :   .::***::           *. :*
```

Figure 5.5. Alanine rich regions within an aligned portion of the N-termini of StzA proteins of Sordariomycetes. This region (aligned using TCoffee) is close to intron 1. Alanine residues are shown in bold (**A**).

5.2.5c Conserved sequences in the C-termini

The C-terminus of *A. nidulans* has been shown to contain a high proportion of aspartic acid (D) and glutamic acid residues (E) (O'Neil *et al*., 2002). With the availability of additional StzA orthologues, it is now clear that these residues consistently occur at a frequency slightly greater than would be expected by chance alone (2 in 20 residues = 10%) in the C-termini of 26 out of 29 StzA orthologues (Table 5.10). The extremely high frequency of proline (P) residues in this region has also previously been observed for *A. nidulans* (O'Neil *et al*., 2002), and it is now evident that this high frequency (approximately 2 to 3 times greater than would be expected by chance: 1 in 20 residues = 5%) is maintained for the C-termini of all StzA orthologues (Table 5.10). Many of these proline residues contribute to SPXX and TPXX motifs, which are commonly found in DNA-binding proteins (Table 5.11; Latchman, 1998), within the C-termini but were also found in frequencies much greater than would be expected by chance in the N-termini. In proteins in general, these motifs only occur at frequencies of 2.89×10^{-3} (SPXX) and 3.08×10^{-3} (TPXX; Dowzer and Kelly, 1991).

Table 5.10. Sizes of C-termini and frequencies of acidic and proline residues within this region. The C-termini were measured from the first residue after the second histidine residue of the third zinc finger to the end of the proteins.

Species	Length of C-terminus (amino acids)	Number of aspartic acid (D) residues	Number of glutamic acid (E) residues	Total frequency of D and E residues	Number of proline (P) residues	Frequency of P-residues
Aspergillus nidulans	198	16	8	12.1%	23	11.6%
Aspergillus fumigatus	302	23	13	11.9%	31	10.3%
Aspergillus oryzae	234	14	11	10.7%	37	15.8%
Aspergillus flavus	234	14	11	10.7%	37	15.8%
Aspergillus niger	251	18	7	10.0%	26	10.4%
Aspergillus terreus	196	14	7	10.7%	22	11.2%
Aspergillus clavatus	302	23	11	11.3%	32	10.6%
Neosartorya fischeri	302	25	12	12.3%	33	10.9%
Ajellomyces capsulatus	363	38	12	13.8%	39	10.7%
Uncinocarpus reesii	147	13	3	10.9%	17	11.6%
Talaromyces emersonii	219	18	9	12.3%	27	12.3%
Coccidioides immitis	276	18	14	11.6%	33	12.0%
Arthroderma benhamiae	339	27	17	13.0%	33	9.7%
Arthroderma gypseum	339	28	16	13.0%	34	10.0%
Trichophyton rubrum	339	26	17	12.7%	33	9.7%
Paracoccidioides brasiliensis	315	29	10	12.4%	29	9.2%
Stagonospora nodorum	255	22	8	11.8%	26	10.2%
Gibberella zeae	252	17	8	9.9%	29	11.5%
Trichoderma reesei	245	19	10	11.8%	27	11.0%
Chaetomium globosum	175	14	5	10.9%	21	12.0%
Neurospora crassa	253	22	8	11.9%	27	10.7%
Botrytis cinerea	223	12	11	10.3%	23	10.3%
Magnaporthe grisea	254	16	8	9.4%	26	10.2%
Podospora anserina	250	21	8	11.6%	26	10.4%
Penicillium chrysogenum	325	20	16	11.1%	34	10.5%
Pyrenophora tritici-repentis	248	21	9	12.1%	27	10.9%
Nectria haematococca	253	18	9	10.7%	30	11.9%
Sordaria macrospora	253	21	8	11.5%	26	10.3%
Verticillium albo-atrum	69	2	4	8.7%	8	11.6%

Table 5.11. Numbers of SPXX and TPXX motifs within the N- and C-termini of StzA proteins.

Species	Number of motifs in the N-terminus		Number of motifs in the C-terminus	
	SPXX	**TPXX**	**SPXX**	**TPXX**
Aspergillus nidulans	4	0	4	3
Aspergillus fumigatus	5	1	6	7
Aspergillus oryzae	2	2	6	3
Aspergillus flavus	2	2	6	3
Aspergillus niger	4	2	12	3
Aspergillus terreus	6	4	4	5
Aspergillus clavatus	6	1	6	5
Neosartorya fischeri	5	1	6	6
Ajellomyces capsulatus	8	2	7	7
Uncinocarpus reesii	1	2	5	3
Talaromyces emersonii	4	5	4	4
Coccidioides immitis	5	1	8	7
Arthroderma benhamiae	7	0	11	6
Arthroderma gypseum	8	0	12	3
Trichophyton rubrum	7	0	12	5
Paracoccidioides brasiliensis	10	2	7	7
Stagonospora nodorum	4	1	5	8
Gibberella zeae	3	1	3	5
Trichoderma reesei	4	3	2	6
Chaetomium globosum	6	3	1	5
Neurospora crassa	3	3	3	5
Botrytis cinerea	3	2	6	3
Magnaporthe grisea	4	5	5	5
Podospora anserina	4	2	2	7
Penicillium chrysogenum	6	1	7	4
Pyrenophora tritici-repentis	4	1	6	8
Nectria haematococca	5	1	3	5
Sordaria macrospora	5	0	3	5
Verticillium albo-atrum	2	2	0	0

Due to the low levels of similarity in the C-termini of StzA proteins (Appendix 6), a more comprehensive study of available databases was undertaken to find any similarities. A search of The Sanger Institute *Schizosaccharomyces pombe* database (http://www.sanger.ac.uk/Projects/S_pombe/) revealed shared similarity between the C-terminus of *A. nidulans* StzA and a region of the *S. pombe* protein encoded by *rsv1* (Figure 5.6). When analysed in detail, 22% amino acid identities were observed, but this included the zinc finger regions for both proteins. When this region was excluded from the comparison and just the low similarity C-terminal region analysed (from *A. nidulans* StzA amino acid 488 to 693), 43 of the 205 amino acids were identical, compared to 46 of these 205 residues for an *A. nidulans–A. oryzae* StzA comparison. Given the generally low level of identity in the C-terminus, it is believed that this similarity may be significant. Intriguingly, the gene *rsv1* encodes a zinc finger TF that is responsive to stress and carbon starvation (Hao *et al.,* 1997).

```
415 KCKDCDKVFARPCDLTKHEKSHS--RPFKCPVTSCKYHIKGWATEKESERHYNDKHSDAP 472   A.nidulans
    +C  C +VF R     +H +SH+  +PF+C   SCK   K +    E  RH    H
  5 ECPFCKRVFHRQEHQVRHIRSHTGEKPFECSYPSCK---KRFTRRDELIRHVRT-H--LR 58    Sch.pombe

473 RLFACQFESCSYKSKRESNCKQHMEKTHGWVYMRSKNNGRSKASPQQQTTSPSSSSVQPK 532   A.nidulans
    +      ++      +  + K   +K+ G    +S+N  +  +       +  + SV
59  KALVTPEQTLDVNLHKAPDSKPEGDKSTGQEADKSQNQSKDGSITDPVQAAVLALSVAYA 118   Sch.pombe

533 QAPSVWSMTPPSEAPDYRQEPNGWDLAPSPETPDLFNTYQAPMTAMPGSVTGTLDAVTPT 592   A.nidulans
    +  SV S++P       +        + S  T  L    Q P+     S++ +  A  PT
119 KPTSV-SLSPTDLQAQSKLIEKPRRRSASNATGSLNKKNQDPLRRF--SISESAGAAAPT 175   Sch.pombe

593 TGTINSPSEPFDLAQENTAFSIQDIFPEMKASDGLLFPGGDMDYPD-----FINNHNMFN 647   A.nidulans
        NS S P     EN    +Q++   + AS+ +  P  + +YP      F++     N
176 PSPSNSKSPP----SENRKNRLQNVLSPI-ASNNV--PDFNQNYPTESNPMFLSTPRFQN 228   Sch.pombe

648 DFGYGDFTMPTQGLQYGETQQPQFEDDSAGFLLDV-YNDMHTYGINP 693               A.nidulans
      G    T+P   +   + +QP     S G  L V YN   +  I P
229 TNGQRTLTVP---VSVWDARQPPTSSRS-GLPLSVMYNHFPSVPIPP 271               Sch.pombe
```

Figure 5.6. Amino acid similarity between the C-terminus of *A. nidulans* StzA and Rsv1 from *Sch. pombe* (Chilton *et al*., 2008).

5.2.6 Phylogenetic analyses of StzA proteins

The sequences derived from manual annotation of StzA proteins were subjected to phylogenetic analyses by using MUSCLE for multiple alignment and PhyML for tree building at Phylogeny.fr (http://www.phylogeny.fr/; Dereeper *et al*., 2008). *Botrytis cinerea* (*Botryotinia fuckeliana*) was chosen as the outgroup because this was the only representative of the Leotiomycetes. Figure 5.7a shows the results for StzA proteins, demonstrating the clustering of species according to the classes to which they belong: Eurotiomycetes, Sordariomycetes and Dothideomycetes. Within these classes, species clustered according to the orders to which they belong: Eurotiales, Onygenales, Sordariales, Hypocreales and Pleosporales. The only peculiarity concerned the clustering of two distinct groups of Onygenales: *A. capsulatus* and *P. brasiliensis* did not cluster with the other members of this class (*A. gypseum*, *A. benhamiae*, *T. rubrum*, *U. reesei* and *C. immitis*).

Since the StzA proteins exhibited great differences in identities in the N- and C-termini compared to the zinc finger regions, the N- and C-termini were themselves subjected to alignments and phylogenetic analyses. When the entire N-terminal region was analysed, a similar tree to that obtained for full-length StzA proteins was observed (data not shown). However, when the C-terminal region was analysed, a very different tree resulted (Figure 5.7b), albeit less robust. The most notable differences concerned *A. terreus*, *A. oryzae* and *A. flavus* grouping with the Sordariomycetes and the isolation of *A. nidulans* and *A. capsulatus* from the other Eurotiomycetes. When the manually annotated *stzA* DNA sequences were subjected to phylogenetic analyses, similar phylogenetic trees were produced (results not shown).

5.7a. Full-length StzA proteins

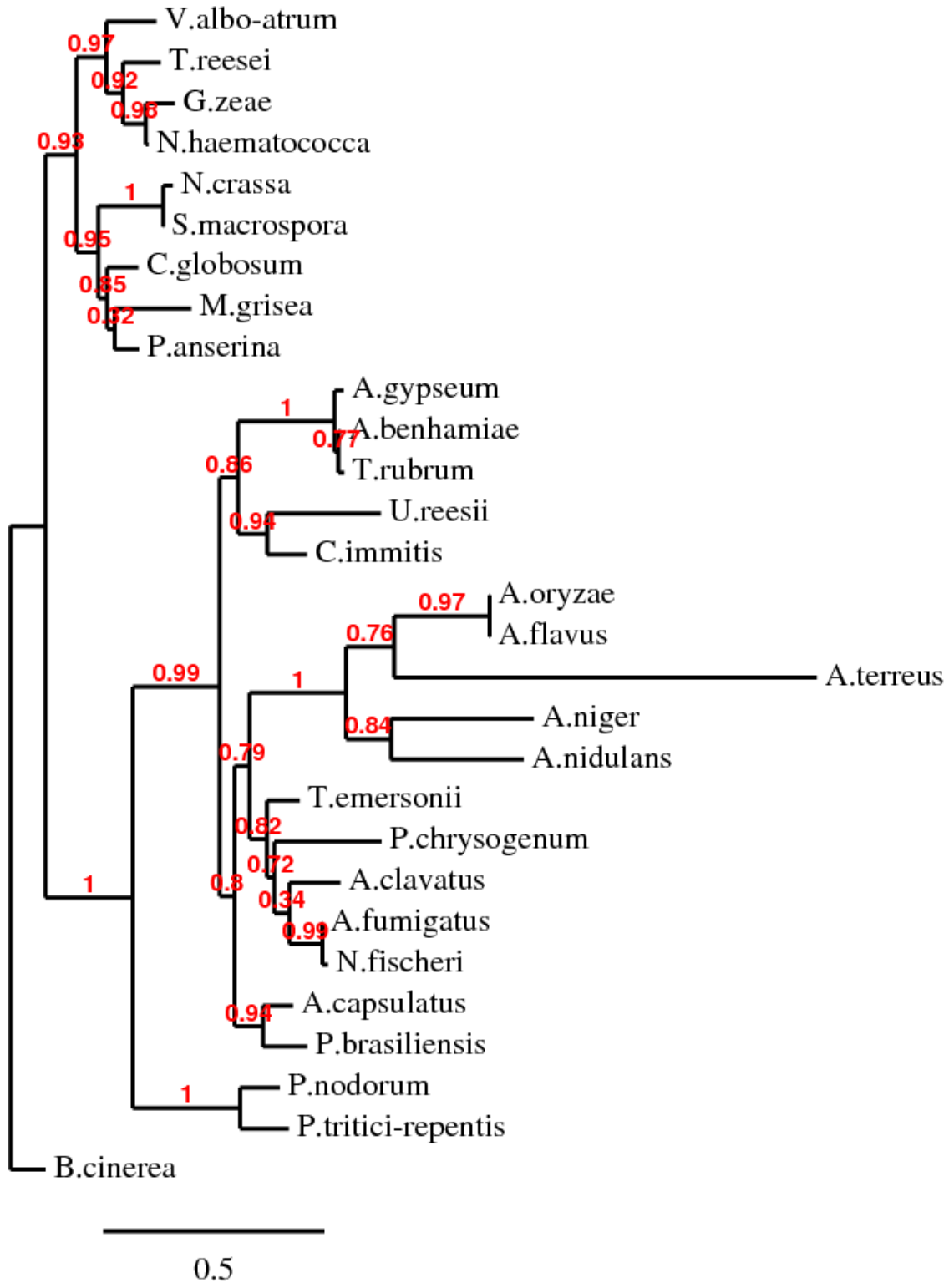


Figure 5.7. Phylogenetic analyses of StzA proteins. Figure 5.7a shows a phylogenetic analysis of full-length StzA proteins, whereas Figure 5.7b shows an analysis restricted to the C-terminal region (continued overleaf). PhyML was used for tree building and TreeDyn for tree drawing using the Phylogeny.fr platform (http://www.phylogeny.fr/; Dereeper *et al*., 2008). 1 = 100% of 100 bootstrap replicates. The scale bar shows the number of substitutions per site. Alternative names for species are shown in Table 5.3.

5.7b. C-termini of StzA proteins

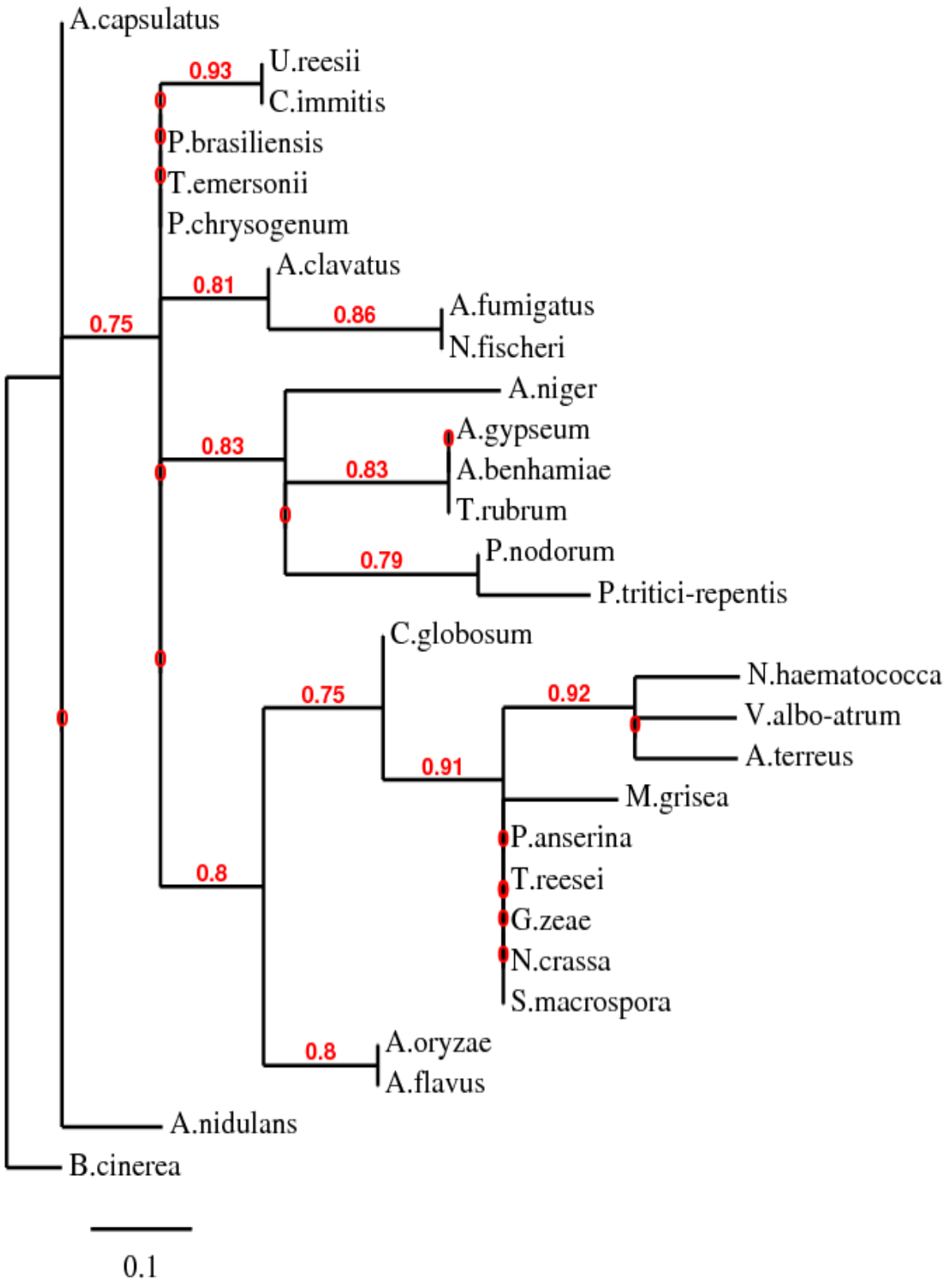


Figure 5.7 (continued).

5.27 Promoter analysis

5.2.7a Identification of potential TF binding sites in *stzA* promoters

A large number of TFs have been reported for *Aspergillus* species together with their experimentally verified DNA binding sequences. These TFs play key roles in regulating fundamental cellular processes, such as metabolism, morphology and development, stress responses and homeostasis (reviewed in Machida and Gomi, 2010). Experimentally verified binding sites for a variety of TFs are shown in Table 5.12 together with their associated regulatory functions. The details provided in the accompanying references are for *Aspergillus* species (usually *A. nidulans*). It is assumed, however, that orthologues of the TFs described are present within other filamentous ascomycetes with *stzA* orthologues, and that these TFs bind to the same target sequences as those described for *Aspergillus* species. For example, it has been experimentally verified that StzA of *A. nidulans* and its orthologue in *T. reesei* (Ace1) both bind to 5'-AGGCA (Aro *et al*., 2003; Spielvogel *et al*., 2008). The positions of potential TF binding sites identified within the promoters (defined as 2 kb regions upstream of the ATG translation start sites) of 22 out of the 29 identified *stzA* orthologues are shown in Table 5.13.

Analysis of the promoter region of *A. nidulans stzA* revealed 4 potential StzA binding sites in its 2 kb promoter region at positions -490, -533, -1192 and -1349 bp upstream of the ATG translation start site. This may indicate that StzA plays a role in its own regulation, although this number of binding site motifs is similar to that expected to occur by chance alone (3.91 / 2 kb). At least one potential StzA binding site was present in all *stzA* promoters except *T. emersonii*, for which only 1067 bp of promoter sequence was available for analysis. In *A. nidulans*, there were also potential binding sites for the following TFs: AbaA (7), AlcR (6), AmdA/AmdX (1), AmyR (3), AnCF (3), AreA (4), BrlA (2), CDRE (4), CpcA (1), CreA (10), PacC (1), PecR (3), STRE (2) and XlnR (1).

Table 5.12. The functions and binding sequences of TFs characterised for *Aspergillus* species.

Transcription factor	Accession number	Function	Binding sequence	Reference
AbaA	AN0422	Conidiophore developmental regulation	5'-CATTCY	Andrianopoulos and Timberlake, 1994.
AflR	AN7820	Aflatoxin regulation	5'-TCGN_5CGA	Fernandes *et al*., 1998
AlcR	AN8978	Ethanol utilisation	5'-CCGCW	Panozzo *et al*., 1997
AmdA / AmdX	AN2270	Acetate and acetamide catabolism	5'-GMGGGG	Andrianopoulos *et al*., 1997 Murphy *et al*., 1997
AmyR	AN2016	Amylolytic transcriptional activation	5'-CGGN_8CGG 5'-CGGAAATTTAA	Petersen *et al*., 1999
AnCF	AN7545 AN4034 AN6492 AN8251	Hap complex Transcriptional regulation	5'-CCAAT	Steidl *et al*., 1999
AreA	AN8667	Nitrogen metabolite repression	5'-HGATAR	Wilson and Arst, 1998
AtfA	AN2911	Conidial stress tolerance	5'-TKACGTMA	Hagiwara *et al*., 2009
BrlA	AN0973	Conidiophore developmental regulation	5'-MRAGGGR	Chang and Timberlake, 1993
CDRE (CrzA)	AN5726	Calcium signalling	5'-GTGGCTG 5'-GAGGCT(G)	Spielvogel *et al*., 2008
CpcA	AN3675	Amino acid biosynthesis	5'-TGACTCA	Hoffmann *et al*., 2001
CreA	AN6195	Carbon catabolite repression	5'-SYGGRG	Cubero and Scazzocchio, 1994
FarA / FarB	AN7050 AN1425	Fatty acid utilisation	5'-CCGAGG	Hynes *et al*., 2006
HacA	AN9397	Endoplasmic reticulum stress	5'-CANRNTGKCCT	Mulder *et al*., 2006
MeaB	AN4900	Nitrogen metabolism	5'-TTGCACCAT	Wong *et al*., 2007
NirA	AN0098	Nitrate assimilation	5'-CTCCGHGG	Strauss *et al*., 1998
PacC	AN2855	pH regulation	5'-GCCARG	Tilburn *et al*., 1995
PecR	AN0741	Pectin regulation	5'-CCCTGA	Coutinho *et al*., 2009
RlmA	AN2984	Cell wall integrity	5'-TA(W_4)TAG	Fujioka *et al*., 2007
STRE	-	Stress response element (*S. cerevisiae*)	5'-AGGGG	Han *et al*., 2004
StuA	AN5836	Developmental regulation	5'-WCGCGWNM	Dutton *et al*., 1997
StzA	AN2919	Stress responses Cellulose utilisation	5'-AGGCA	Saloheimo *et al*., 2000 Spielvogel *et al*., 2008
XlnR	AN7610	Xylanolytic and celluloytic utilisation	5'-GGCTAR	De Vries *et al*., 2002

Table 5.13. Potential binding sites for a range of TFs within *stzA* promoters (continued overleaf). *stzA* promoter sequences of 2 kb were manually analysed for potential TF binding sites (forward and reverse complements) using the Microsoft Word® "Find" tool. The *stzA* promoters of *T. emersonii* and *G. zeae* had only 1067 bp and 589 bp available for analysis. In the promoter of the *B. fuckeliana stzA* orthologue, nucleotides 261–620 were denoted with an 'N' so could not be analysed. All numbers represent the position of motifs upstream of the ATG translation start sites. Orthologues that contain no binding site motifs for a particular TF are shown by the symbol (-). The accession numbers, functions and binding sequences of the TFs are listed in Table 5.12.

Transcription factor	Species																				
	A. nidulans	*A. fumigatus*	*N. fischeri*	*A. clavatus*	*A. oryzae*	*A. flavus*	*A. niger*	*A. capsulatus*	*A. terreus*	*T. emersonii*	*C. immitis*	*U. reesii*	*P. nodorum*	*P. tritici-repentis*	*T. reesei*	*M. grisea*	*C. globosum*	*N. crassa*	*G. zeae*	*P. anserina*	*B. fuckeliana*	*P. chrysogenum*
AbaA	-168 -448 -1107 -1282 -1409 -1541 -1702	-1222 -1493 -1706	-1192 -1258 -1316 -1696	-281 -738 -1292 -1494	-68 -690 -696 -1362 -1378	-68 -690 -696 -1362 -1378	-446	-330 -562 -682	-605	-408 -439	-854 -1924	-1191 -1402 -1482	-196	-1772	-	-94 -578 -1780	-1073 -1960	-810 -1557 -1884	-158	-705 -1832 -1987	-106 -1084	-602 -705 -724 -1802
AflR	-	-	-	-	-	-	-340	-	-293 -884	-	-	-	-	-	-	-184	-133	-	-	-	-1037	-
AlcR	-481 -719 -912 -1121 -1205 -1753	-805 -1027 -1134	-50 -476 -820	-49 -86 -368 -873 -888 -896 -1070 -1424 -1463 -1471 -1721 -1971	-264 -593 -1521 -1556 -1756 -1791	-264 -593 -1521 -1556 -1756 -1791	-39 -118 -160 -530 -857 -1145 -1258 -1288 -1719 -1776	-199 -242 -782 -866 -885 -982 -1070 -1180 -1300 -1447 -1489 -1522 -1545 -1782 -1960 -1980	-97 -217 -236 -351 -1481 -1789 -1877	-33 -41 -123 -178 -343 -391	-64 -208 -313 -324 -531 -780 -1382	-705 -782 -1164 -1197 -1227 -1294 -1398	-116 -156 -315 -395 -418 -657 -690 -743 -792 -929 -958 -1082 -1249 -1327 -1461 -1539 -1767	-121 -293 -323 -663 -861 -894 -1329 -1769 -1777	-80 -87 -666 -733 -790 -824 -1002 -1053 -1233 -1318 -1486 -1581 -1724 -1895	-17 -127 -900 -915 -1007 -1018 -1043 -1485 -1892	-107 -188 -450 -611 -809 -1153 -1491 -1596 -1768 -1927	-131 -868 -1091 -1123 -1191 -1251 -1786	-	-4 -13 -598 -1047 -1073 -1519 -1677 -1836 -1844 -1879	-809 -968 -1056 -1761	-67 -199 -353 -616 -825 -1308 -1608 -1736
AmdA/AmdX	-717	-	-	-696 -1673	-149 -478 -918	-149 -478 -918	-871	-17 -864 -956 -1178	-1247 -1479	-	-1826	-	-256 -357 -426 -658 -1414	-341 -1253	-976	-18 -128 -837 -1073 -1890	-15 -1068 -1602 -1609	-89 -892 -1026	-	-14 -312 -1079 -1197	-903 -1643 -1692 -1925	-332 -354 -810 -1443
AmyR	-215 -590 -811	-1134 -1239	-	-1452 -1471	-	-	-109	-733 -1687 -1869	-488	-8 -43 -569	-315	-	-306 -1873	-43 -54 -284 -1263	-1242 -1243	-1007	-1180 -1181	-	-	-600	-	-1617
AnCF	-138 -738 -1099	-65 -1374	-65 -653 -1370	-568 -989	-496 -530 -605 -647 -777 -872 -1278 -1311 -1448 -1577 -1867	-496 -530 -605 -647 -777 -872 -1278 -1311 -1448 -1577 -1867	-704 -1268 -1470 -1509 -1516 -1627 -1703 -1740	-132 -292 -348 -872 -1133	-510 -636 -706 -1369 -1474 -1778	-	-44 -229 -274 -762 -791 -820 -1313 -1703 -1949	-42 -770 -714 -1537	-604 -1442 -1589	-179 -563 -1087 -1113 -1290 -1554 -1746 -1909 -1940	-297 -397 -571	-211 -287 -448 -593 -621 -689 -910 -1391 -1530 -1877	-711 -732 -744	-208 -480 -899 -993 -1219 -1366 -1668	-248 -410	-213 -218 -899 -921 -945 -964 -973 -1787 -1943	-676 -826	-2 -75 -460 -740 -770 -1153 -1239 -1330 -1592 -1944
AreA	-537 -1077 -1165 -1510	-358 -495 -689 -729 -835	-512 -528 -644 -703 -743 -850	-364 -417 -646 -744 -757 -777 -929 -952 -1045 -1901	-57 -209 -470 -525 -1010 -1418 -1514 -1705 -1848	-57 -209 -470 -525 -1010 -1418 -1514 -1705 -1848	-535 -646 -843 -1167 -1744 -1767 -1971	-408 -846 -1079 -1086 -1242 -1255	-230 -301 -751 -1334 -1593	-835 -926	-1710 -1985	-451 -473 -688 -1684 -1696 -1782	-187 -716 -969 -1134 -1223 -1821	-161 -168 -945 -1222 -1371 -1838	-1589 -1697	-115 -405 -485 -1793 -1804	-244 -271 -484 -759 -1476 -1585	-185 -266 -388 -1983	-292 -337	-48 -186 -279 -655 -892 -996 -1590 -1896	-126 -631 -643 -1256 -1264 -1594 -1706	-170 -270 -276 -364 -672 -1257 -1582
AtfA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
BrlA	-329 -574	-342 -860 -1170	-341 -733 -875 -1380 -1619	-520 -667 -1305 -1564 -1597	-241 -477 -1946	-241 -477 -1946	-247 -308	-116 -775	-1247	-252	-266 -672 -1290 -1376 -1721	-	-356 -425 -1414 -1849	-340 -381	-98 -647	-768 -836 -1073	-526 -1328 -1644	-1026 -1261 -1692	-	-301 -505 -1079 -1197 -1237 -1341	-1159 -1643 -1696	-810 -1443
CDRE (CrzA)	-510 -1063 -1317 -1448	-303 -911	-789 -926	-1757 -1988	-443 -1504 -1976	-443 -1504 -1976	-606 -696 -1577 -1640 -1934	-1335 -1950	-430 -473 -1194	-	-877 -1342 -1523 -1716	-1154	-367 -495 -1273	-	-127 -1845	-64	-455 -1338	-1107	-	-1717	-	-644 -909 -1296 -1359
CpcA	-67	-107	-107	-91	-1152 -1586	-1152 -1586	-96	-139	-1362	-87 -151	-79	-87	-55	-58	-169 -306	-139	-123	-119	-144	-130	-93	-89
CreA	-608 -717 -819 -851 -1356 -1451 -1598 -1826 -1953 -1981	-927 -1000	-942 -1022 -1757	-105 -234 -696 -1425 -1972	-553 -739 -822 -918 -1789 -1893	-553 -739 -822 -918 -1789 -1893	-40 -268 -300 -388 -531 -855 -871 -923 -996 -1092 -1256 -1279 -1286 -1447	-102 -214 -434 -607 -864 -1178 -1323 -1426 -1560 -1570 -1582 -1597 -1723 -1944	-26 -535 -619 -793 -896 -935 -1158 -1354 -1479 -1695 -1875	-119 -287 -350 -612 -810	-325 -826 -927 -929 -1195 -1370 -1410 -1443 -1599	-187 -347 -725 -1222 -1964	-171 -256 -337 -653 -658 -1119 -1239 -1425 -1899	-146 -484 -594 -602 -822 -917 -1144 -1253 -1261 -1592	-35 -468 -524 -667 -675 -740 -976 -1003 -1023 -1166 -1178 -1418 -1434 -1826 -1841	-18 -29 -128 -544 -889 -923 -941 -1476 -1890	-15 -108 -179 -434 -451 -650 -924 -963 -1068 -1246 -1467 -1492 -1609	-15 -222 -447 -638 -781 -878 -892 -1246 -1576 -1701 -1963	-	-14 -271 -635 -810 -1611 -1663 -1707 -1754	-64 -1850 -1925	-96 -332 -354 -436 -503 -639 -845 -984 -1023 -1451 -1737 -1764
FarA/FarB	-	-326	-988	-	-1948	-1948	-1004	-	-	-336	-423	-	-583	-	-229	-767	-	-104	-380	-	-	-
HacA	-	-1492	-	-	-	-	-	-	-	-	-	-	-	-	-	-101	-	-	-	-	-	-
MeaB	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NirA	-	-	-	-	-	-	-1004	-	-	-	-	-	-169	-	-	-	-	-	-	-	-	-

Table 5.13 (continued).

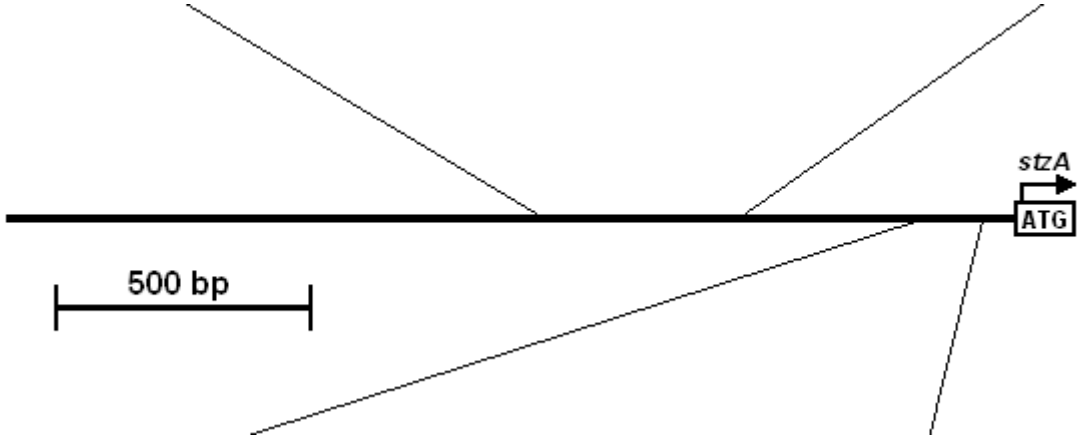
Transcription factor	Species																			
	A. nidulans	*A. fumigatus*	*N. fischeri*	*A. clavatus*	*A. oryzae*	*A. flavus*	*A. niger*	*A. capsulatus*	*A. terreus*	*T. emersonii*	*C. immitis*	*U. reesii*	*P. nodorum*	*P. tritici-repentis*	*T. reesei*	*M. grisea*	*C. globosum*	*N. crassa*	*G. zeae*	*P. anserina*	*B. fuckeliana*	*P. chrysogenum*
PacC	-1331	-650 -972 -1974	-1035 -1964	-	-561 -816 -1292	-561 -816 -1292	-755 -1396	-29 -283 -616 -856 -1743 -1993	-1180	-23 -424	-708 -1334 -1768	-1089 -1504	-252 -520	-968 -1385 -1801	-1083 -1281 -1934	-1 -387 -393 -1123 -1398	-294 -677 -715	-212 -304 -559 -1509	-121 -218	-364 -398 -589 -854	-841	-1145
PecR	-282 -467 -1484	-1087	-53 -1101	-1296 -1487	-	-	-1333	-	-941	-296	-587 -1328	-	-	-1228	-	-	-	-	-	-908 -1402	-1939	-260
RlmA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-1272 -1367	-
STRE	-331 -576	-847 -1089 -1172 -1759	-862 -1382 -1619	-1307 -1597 -1674	-150 -243 -479 -1744	-150 -243 -479 -1744	-81 -249 -310 -371 -421 -1675 -1888	-18 -118 -415 -714 -957	-1247	-130 -254 -295 -315 -395	-683 -1290 -1376 -1798 -1826	-639	-358 -427 -1414 -1849	-342 -583 -1834	-139 -750	-838 -1073	-675 -1546 -1646 -1900	-90 -182 -1026	-178	-308 -313 -484 -507 -743 -1079 -1197 -1343 -1404	-904 -1643 -1692	-810 -1443
StuA	-	-	-	-	-	-	-	-	-142 -576 -578	-	-	-245 -247	-529 -1061 -1468	-631 -633	-406	-1901 -1903	-577 -579 -976 -978 -1413	-1580	-	-733 -735	-769 -771	-
StzA	-490 -533 -1192 -1349	-9 -350 -831 -1015 -1048 -1388 -1946	-9 -349 -846 -1936	-8 -510 -925 -1360	-309 -340 -466 -1592	-309 -340 -466 -1592	-57 -642 -1478	-842 -1103 -1889	-93 -456 -611 -1022 -1263	-	-488 -577 -731	-333 -455 -684	-513 -1482 -1556 -1577 -1652	-1517 -1936	-546 -945	-382 -876 -1480	-635 -1020	-196 -666 -835 -972 -1831	-186	-831 -874 -1181	-1073	-541 -668 -926 -1051 -1893 -1952
XlnR	-1258	-88 -1230	-1234	-195	-879 -1807	-879 -1807	-662 -1936	-	-475	-427	-	-1544	-1124 -1170 -1699	-1558 -1711	-	-1218	-891	-562 -722	-	-	-58 -1019	-

Within the promoter of *A. nidulans stzA*, it was noted that there are overlapping canonical AreA (5'-HGATAR; Wilson and Arst, 1998) and StzA (5'-AGGCA; Spielvogel *et al*., 2008) binding site motifs at -533 bp upstream of the ATG translation start site and comprise the sequence 5'-CTATCAGGCA. This would appear to acquire even greater significance considering that the exact conservation of this sequence is retained in the promoters of 10 other *stzA* orthologues: *A. fumigatus* (-831), *A. oryzae* (-466), *A. flavus* (-466), *A. niger* (-642), *A. clavatus* (-925), *N. fischeri* (-846), *A. capsulatus* (-842), *U. reesii* (-684), *P. chrysogenum* (-668) and *P. brasiliensis* (-975). Figure 5.8a shows that the sequence 5'-CTATCAGGCA is part of the larger sequence TCG(T/C)CTATCAGGCAGAC conserved in the promoters of all eleven *stzA* orthologues belonging to the Eurotiomycetes class of the Pezizomycotina subphylum. Apart from this region, surprisingly little nucleotide conservation was evident in *stzA* promoters when three alignment programs were used (ClustalW, TCoffee and Mlagan), even when alignments were restricted to *Aspergillus* species. A search of the entire genome of *A. nidulans* for the sequence 5'-CTATCAGGCA, containing putative overlapping AreA/StzA binding sites, was undertaken. This sequence was identified in the promoters of several genes, including the *pol* region of a retrotransposon that is related to the *REAL* LTR retrotransposon of *Alternaria alternata* (Kaneko *et al*., 2000). Hence, the sequence was termed *REAL* associated little element (*REALALE*).

It was evident that 26 of the 29 *stzA* promoters contained a CpcA binding site motif (5'-TGACTCA; Hoffmann *et al*., 2001) within a similar region (-55 to -169 bp from the ATG translation start site; Figure 5.8b). The exceptions, *A. oryzae*, *A. flavus* and *A. terreus*, contained no CpcA binding site motifs in this region of their promoters, although the closely matching sequence TGACTAC (-62) was identified in *A. oryzae* and *A. flavus*, and TGAGTGA (-83) in *A. terreus*. Canonical CpcA binding sites motifs were, however, present in more upstream promoter regions for *A. oryzae* and *A. flavus* (-1152 and -1586 for both species) and *A. terreus* (-1362).

(a)

```
A.nidulans      -CCTTTTAGGGATTAGCACATCTCGTCTATCAGGCAGACGTC-CAAG-ACGCCTCAGCCT  -533
A.oryzae        -CCAATTGTGGCTTAGATCCCCTCGTCTATCAGGCAGACATC-CAAG-ACGACTCAGCCT  -466
A.flavus        -CCAATTGTGGCTTAGATCCCCTCGTCTATCAGGCAGACATC-CAAG-ACGACTCAGCCT  -466
A.fumigatus     CCCTTTCTGGGATTAGGG-GGCTCGCCTATCAGGCAGACATC-GAAA-CCACCTCAGCTT  -831
N.fischeri      CCCTTTCTGGGATTAGGG-GGCTCGTCTATCAGGCAGACATC-GAAA-CCACCTCAGCTT  -846
A.clavatus      CCCTATCTCAGCCTAGCGTG-CTCGTCTATCAGGCAGACACCTGAAC-GTCGTCGAAGCC  -925
A.niger         -CCTAGAGTAGGCTAGCGCAACTCGTCTATCAGGCAGACATC-GAAG-ACGACTCAGCCC  -642
U.reesii        -CCAATGACGAGCGGCGTCCGTTCGTCTATCAGGCAGACATCAAAAGCACGTCTCGGCGT  -684
A.capsulatus    -CCAATCAGCGGGGGCCCTGGCTCGCCTATCAGGCAGACACG-ATGCTTCGCCTTGGTAC  -842
P.chrysogenum   -CCATTTCGGTGTTAGCGTA-TTCGTCTATCAGGCAGACATC-CAAC-ACGCAATGAGCC  -668
P.brasiliensis  -CCAATCAGGAGGCTCCCTTGCTCGTCTATCAGGCAGACACGATTGCTACGACTCCCCTT  -975
                 **                    *** *************
```



(b)

```
A.nidulans          -----GATTTGGTCACGATCAGTGAGTCATTT---ACCTGCATCA---CTCT----    -67
A.fumigatus         ----TATTACATCGCACC-AGGTGAGTCACAG-----ACAGGGT--GGTCTAG---   -107
N.fischeri          ----TATTACATCGCACC-AGGTGAGTCACAG-----ACAGGGT--GGTCCAG---   -107
A.clavatus          -----ATTACTCCGCGTGTGAGTGAGTCACC---GCAGCTGGTCA---GCAT----    -91
A.niger             ---CAAACATACCGGAC--AAGTGAGTCAC--------CGCGGC--AATAGGGGTG    -96
T.emersonii         ----TCGCCCTCTGCATG-TGGTGAGTCACCCTCGTGGCCGATT-----CC-----    -87
P.chrysogenum       ---TCTGGACAACTCTCC--GGTGAGTCATGA-----GATGGGCC-AATCAG----    -89
U.reesii            -----TCGACATTGAACGAAGGTGAGTCACAG-----ATGGATCT-AATTCG----    -87
A.capsulatus        ---CACTTTTGCTTTCCA--GGTGAGTCATA-------CCAATCTATGTCCTC---   -139
C.immitis           --GCTTCCACTCAGACTC---GTGAGTCATAA-----ACGGATC-----CCGCAAA    -79
A.gypseum           CCGAAAGATT-AAAGAG----GTGAGTCATA-------TGAAGCT--ATTCCCTT-    -91
A.benhamiae         CCGAAAGAAT-CAGCTG----GTGAGTCATA-------TGGAACT--ATCCTTTT-    -90
T.rubrum            CCGAAAGAAT-CAGCTG----GTGAGTCATA-------TGGAACT--ATCCTTTT-    -90
P.brasiliensis      -----TCCCCTCTGTTTTCCAGTGAGTCATA---ACC-CTAGTTT---ACAAC---   -140
P.nodorum           ---AAAGTCTACAGGTGC--AGTGACTCATAGC----CTGAGGCC--TTCCA----    -55
P.tritici-repentis  -GGGAACTGTGCAGTCC----GTGACTCA--------TCCGACCTGAGGCCGCC--    -58
C.globosum          -GTGTCGCAGGCCGACT----GTGACTCACGC-----TCTCATCT----CCGCTC-   -123
N.crassa            -----CATACTGCCGCTTCCTGTGACTCAC-A---CGAC--GTTC---CTCGGTC-   -119
P.anserina          -----GGAATATCCTGCTTCTGTGACTCACAT---CCTCTCGTCT---CTCA----   -130
S.macrospora        -----CATACTTCCACTCCCTGTGACTCAC-T---CGAC--GTTC----TCGGTCT   -119
M.grisea            ----CCGCGCT-TGCGTGGTTGTGACTCACAT--ATCCCCGCTT-----CAGC---   -139
T.reesei            ---CGTCTATCTCGGGC--TTGTGACTCACAC---CTGT-CGACC---CCCCC---   -169
G.zeae              -GTTAGGAATGCCGA----CTGTGACTCACAT---CTCC--GTCC---TTCACT--   -144
N.haematococca      -ATTAGGAACCTCGA----CTGTGACTCACCA---TCCC--GTCC---TCCACT--   -146
V.albo-atrum        --CCAGACTTCTCAGG---CAGTGACTCACTT---CCCGACACCT---TCCT----   -127
B.cinerea           --CTATTCAT-TCTCAT--TTGTGACTCACAT---TTCT-TGCAT---GGAAT---    -93
                                       **** ***
```

Figure 5.8. ClustalW alignments showing overlapping canonical AreA/StzA binding site motifs (*REALALE* sequences) and CpcA binding site motifs in *stzA* promoters. (a) The *REALALE* sequence 5'-CTATCAGGCA, which contains putative AreA (CTATCA - blue) and StzA (AGGCA - red) binding sites that overlap (purple) is shown underlined and is present in eleven *stzA* promoters, including seven *Aspergillus* species. Numbers show the positions of the sequences upstream of the ATG translation start sites. Note that the *REALALE* sequence is part of the conserved sequence TCG(T/C)CTATCAGGCAGAC. (b) A putative CpcA binding site (TGA(G/C)TCA) is present and conserved in position in the vast majority of *stzA* promoters (26 out of 29) and is shown underlined and highlighted in green. The putative CpcA binding sites are part of the sequence GTGA(G/C)TCA(C/T). Eurotiomycetes share the same putative CpcA binding sequence 5'-TGAGTCA, whereas the Sordariomycetes, Dothideomycetes and the one representative Leotiomycete share the sequence 5'-TGACTCA.

Chi-square analysis was used to assess whether promoters of *stzA* genes were significantly enriched for the presence of specific TF binding site motifs when compared to a control group of genes. Coutinho *et al*. (2009) used a similar approach to determine the percentage of promoters of genes from functionally related groups in *Aspergillus species* that were significantly enriched for binding sites of TFs involved in plant polysaccharide degradation. Enrichment of promoters for TF binding motifs indicates an increased likelihood of functional significance (i.e. that a specific TF actually binds to some of the motifs in a promoter to regulate expression of the gene). Assuming equal occurrences of the nucleotides A, T, C and G in fungal promoters, the occurrences of random sequences for specific TF binding sites could be deduced. Based on these frequencies, chi-square analysis could then be used to calculate the minimum number of motifs required to demonstrate a specific enrichment of that motif in a 2 kb promoter (Table 5.14).

Table 5.14. Predicted random occurrences of TF binding site motifs and their frequencies required to demonstrate significant enrichment. Equal occurrences of the bases A, T, C and G were assumed when predicting the occurrences of TF binding site motifs (shown in Table 5.12) in promoters by random chance. The minimum numbers of motifs required to be present in promoters (2 kb) to demonstrate significant enrichment according to chi-square analysis are shown. The critical value of chi-square is 3.84 for one degree of freedom at the 5% significance level (O = observed, E = expected).

Transcription factor	Predicted occurrence of binding motif by chance	Significantly enriched cut-off value	Chi-square calculation $(O - E)^2 / E$
AlcR CreA	1 in 256 bp (7.81 / 2 kb)	14	When O = 13, χ^2 = 3.45 When O = 14, χ^2 = 4.91
AreA	1 in 384 bp (5.21 / 2 kb)	10	When O = 9, χ^2 = 2.76 When O = 10, χ^2 = 4.40
AnCF STRE StzA	1 in 512 bp (3.91 / 2 kb)	8	When O = 7, χ^2 = 2.44 When O = 8, χ^2 = 4.28
AbaA AmdA / AmdX BrlA CDRE (CrzA) PacC StuA XlnR	1 in 1,024 bp (1.95 / 2 kb)	5	When O = 4, χ^2 = 2.16 When O = 5, χ^2 = 4.77
AflR AmyR FarA / FarB PecR RlmA	1 in 2,048 bp (0.98 / 2 kb)	3	When O = 2, χ^2 = 1.06 When O = 3, χ^2 = 4.16
AtfA CpcA	1 in 8,192 bp (0.24 / 2 kb)	2	When O = 1, χ^2 = 2.41 When O = 2, χ^2 = 12.91
NirA	1 in 12,288 bp (0.16 / 2 kb)	1	When O = 1, χ^2 = 4.41
HacA	1 in 32,768 bp (0.061 / 2 kb)	1	When O = 1, χ^2 = 14.45
MeaB	1 in 131,072 bp (0.015 / 2 kb)	1	When O = 1, χ^2 = 64.68

The percentage of *stzA* promoters (N = 19) demonstrating enrichment for specific TFs is shown in Table 5.15, which also shows comparisons with results obtained using a control group of gene promoters (N = 22; Appendix 8). Numbers of *stzA* promoters demonstrating enrichment for TF binding site motifs were generally low (<25%) and similar to those found in the control group of promoters. The *A. nidulans stzA* promoter is enriched for the TFs AbaA, AmyR and PecR, which control conidiophore development, amylose utilisation and pectin utilisation, respectively (Appendix 9; Andrianopoulos and Timberlake, 1994; Petersen *et al*., 1999; Coutinho *et al*., 2009). Eight of the 19 *stzA* promoters (42.1%) analysed were enriched for the AnCF binding site motif 5'-CCAAT (Steidl *et al*., 1999) compared to 4 promoters (18.2%) in the control group. In *A. nidulans*, AnCF serves as a redox sensor in co-ordinating the cellular oxidative stress response (Thön *et al*., 2010). The data indicates that it is unlikely that the TFs AflR, AreA, CDRE, FarA, FarB, HacA, MeaB, NirA, PacC, PecR, STRE, StuA, StzA or XlnR contribute to the regulation of *stzA* when considering the group of *stzA* promoters as a whole.

Table 5.15. Percentages of *stzA* promoters significantly enriched for specific TF binding motifs compared to a control group of promoters. Nineteen *stzA* promoters were used in the analysis (all *stzA* promoters in Table 5.13 except those of *T. emersonii*, *G. zeae* and *B. fuckeliana*, for which complete promoters were not available). The control group comprised 22 genes randomly selected from the *A. nidulans* genome (Appendix 8). The complete data sets are shown in Appendix 9.

Transcription factor	Percentage of promoters significantly enriched for the TF binding motif	
	***stzA* promoters**	**Control promoters**
AbaA	15.8%	9.1%
AflR	0.0%	0.0%
AlcR	15.8%	0.0%
AmdA / AmdX	10.5%	4.5%
AmyR	15.8%	4.5%
AnCF	42.1%	18.2%
AreA	5.3%	0.0%
BrlA	21.1%	4.5%
CDRE	5.3%	4.5%
CpcA	15.8%	0.0%
CreA	21.1%	13.6%
FarA / FarB	0.0%	9.1%
HacA	10.5%	9.1%
PacC	10.5%	18.2%
PecR	5.3%	9.1%
STRE	5.3%	4.5%
StuA	5.3%	0.0%
StzA	0.0%	9.1%
NirA	10.5%	22.7%
MeaA	0.0%	4.5%
XlnR	0.0%	0.0%

Three *stzA* promoters were enriched for the CpcA binding site motif (5'-TGACTCA; Hoffmann *et al*., 2001) compared to none in the control group. However, this percentage (15.8%) appears to underestimate the significance of CpcA in contributing to *stzA* expression, considering that every *stzA* promoter contains at least one CpcA binding site motif. Furthermore, 26 of these motifs are present in a conserved position within 200 bp from the ATG translation start site (Figure 5.8b), and yet by chance alone, a CpcA binding site motif is expected to occur only once in every 8,192 bp. Hence, this conservation is indicative of a functional role for the binding of CpcA, which mediates the cross pathway control of amino acid biosynthesis, to *stzA* promoters (Hoffmann *et al*., 2001). In the control group, it would be expected that approximately 1 in 4 promoters possess a CpcA binding site motif by chance alone. This is similar to the actual percentage of promoters possessing a CpcA binding site motif (31%), and is reasonable considering a couple of these control genes could actually be regulated by CpcA, and therefore may possess more CpcA binding sites than would be expected by chance alone.

5.2.7b Identification of the *REALALE* sequence in the promoters of *Aspergillus* genes

There were a total of 67 occurrences of the *REALALE* sequence (5'-CTATCAGGCA) within the *A. nidulans* genome. The *REALALE* sequence occurred within the 2 kb promoter of 46 predicted genes and on a further 21 occasions either within the ORF of a gene (14 occasions) or within intergenic regions (7 occasions). The sequence was associated with an LTR retrotransposon on 13 occasions (on 6 occasions within promoters and on 7 occasions within ORFs). The functions (known or putative) of genes containing the *REALALE* sequence within their promoters were identified at the Broad Institute (www.broadinstitute.org/) and the Central *Aspergillus* Data REpository (CADRE; http://www.cadre-genomes.org.uk/) and are shown in Table 5.16. Orthologues of *A. nidulans* genes containing the *REALALE* sequence in *A. fumigatus* and *A. oryzae* were identified using the proteins for BLAST analysis at NCBI (www.ncbi.nlm.nih.gov/). The corresponding promoters were subsequently analysed for the *REALALE* sequence. No orthologues contained this sequence within their promoters, with the sole exception of *stzA*.

Table 5.16. *A. nidulans* gene promoters containing the *REALALE* sequence (continued overleaf). Promoters of genes containing either the forward complement (CTATCAGGCA) or the reverse complement (TGCCTGATAG) of the *REALALE* sequence are shown. The position of each *REALALE* sequence upstream of the ATG translation start is indicated. Numbers of potential StzA binding sites within each 2 kb promoter are also shown. Orthologues of these genes were identified in *A. fumigatus* and *A. oryzae*, and their promoters were subsequently analysed for the presence of the *REALALE* sequence, which did not occur in any promoter apart from that of *stzA*.

Accession number	Known or predicted protein function	Region of promoter containing the *REALALE* sequence	No. of potential StzA binding sites (2 kb promoter)	Orthologues in *A. fumigatus* and / or *A. oryzae*
AN0539	Hypothetical protein (contains tetratricopeptide repeat).	-1456	9	AO090103000078
AN1731	Proline oxidase (PrnD).	-1381	6	AFUA_6G08760 AO090001000550
AN1732	Proline-specific permease (PrnB).	-274	7	AFUA_8G02200 AO090010000119
AN2424	Beta-N-acetylhexosaminidase. (Glycosyl hydrolase family 20).	-513	3	AFUA_2G00640 AO090701000314
AN2978	Hypothetical protein. No protein domains found.	-299	9	AFUA_3G08440
AN5188	Ribonuclease H.	-1481	7	AFUA_1G07180 AO090012000971
AN5189	Hypothetical protein. No protein domains found.	-116	9	
AN5924	Putative $Zn(II)_2Cys_6$ TF.	-790	7	AFUA_2G10850
AN6104	Hypothetical protein. No protein domains found.	-233	9	AFUA_2G09510 AO090011000722
AN6498	Hypothetical protein.	-1715	4	AFUA_6G05240 AO090701000011
AN6966	LTR-retrotransposon.	-290	10	AO090005001597
AN7202	Isochorismatase family hydrolase.	-650	5	AFUA_3G03270 AO090001000095
AN8295	LTR-retrotransposon.	-290	10	
AN8305	LTR-retrotransposon.	-290	Incomplete promoter	AO090005001597
AN8982	Fructosyl amine:oxygen oxidoreductase.	-1243	1	AFUA_2G02030 AO090011000473
AN9220	Toxin biosynthesis protein (GliH).	-1846	3	AFUA_5G10320 AO090026000716
AN9294	Hypothetical protein. (Major facilitator superfamily).	-136	4	AFUA_4G03750 AO090011000352
AN9312	Hypothetical protein. No protein domains found.	-1131	4	
AN1079	Arginine rich protein.	-1003	7	AFUA_1G12120 AO090001000333
AN1302	Hypothetical protein. No protein domains found.	-1922	6	AFUA_8G00600 AO090102000193
AN1893	Putative $Zn(II)_2Cys_6$ TF.	-1208	6	AFUA_2G04262 AO090003000213
AN1894	Hypothetical protein. No protein domains found.	-822	7	
AN1949	ATP-dependent RNA helicase (Has1).	-158	5	AFUA_4G13330 AO090009000242
AN2028	LTR-retrotransposon.	-678	9	
AN2919	StzA. C_2H_2 finger domain TF related to Ace1.	-533	4	AFUA_3G08010 AO090005001502
AN2920	Hypothetical protein. No protein domains found.	-319	7	
AN2951	UDP-glucose 4-epimerase.	-1295	6	AFUA_3G07910 AO090005001490
AN3000	Hypothetical protein. No protein domains found.	-1088	3	AFUA_3G08690 AO090005001398
AN3585	Hypothetical protein. No protein domains found.	-522	7	AFUA_4G12940 AO090009000293
AN3586	Ubiquinone biosynthesis monooxgenase (Coq6).	-584	8	AFUA_4G12930 AO090009000294
AN4013	C_2H_2 TF.	-1408	3	AFUA_1G04110 AO090003000941
AN4924	Hypothetical protein. No protein domains found.	-1660	8	
AN4925	Glutamate carboxypeptidase (Tre2).	-402	8	AFUA_3G10650 AO090003000609

Table 5.16 (continued).

Accession number	Known or predicted protein function	Region of promoter containing the *REALALE* sequence	No. of potential StzA binding sites (2 kb promoter)	Orthologues in *A. fumigatus* and / or *A. oryzae*
AN5471	LTR-retrotransposon.	-290	9	AO090005001597
AN5606	Vacuolar ATPase 98 kDa subunit.	-1400	6	AFUA_4G11300 AO090003001090
AN6705	Component of the RSC chromatin remodeling complex. Orthologue of *S. cerevisiae RSC8* involved in double-strand break repair *via* nonhomologous end-joining.	-846	6	AFUA_7G05510 AO090005000416
AN7262	Hypothetical protein. Contains a repeat of unknown function.	-1865	6	AFUA_2G17000 AO090102000117
AN7642	Aconitate hydratase.	-299	7	AFUA_6G12930 AO090001000642
AN7486	LTR-retrotransposon.	-682	9	
AN8114	Hypothetical protein. No protein domains found.	-191	9	AFUA_6G09040 AO090010000573
AN8188	GTP cyclohydrolase I.	-156	2	AFUA_5G03140 AO090102000517
AN8189	Hypothetical protein. No protein domains found.	-1635	4	AFUA_5G03150 AO090102000518
AN8479	Hypothetical protein. No protein domains found.	-684	5	AFUA_1G13230 AO090009000073
AN8902	Putative $Zn(II)_2Cys_6$ TF.	-1477	2	AFUA_2G03430 AO090026000279
AN8939	PHD finger protein.	-549	7	AFUA_6G13765
AN9308	FAD binding domain protein.	-64	5	AFUA_2G00730 AO090701000319

A promoter containing *REALALE* is one of the best-studied promoter regions of *A. nidulans*, namely the *prnD–prnB* intergenic *cis*-acting regulatory region of the proline utilisation cluster (Sophianopoulou and Scazzocchio, 1989; Gómez *et al*., 2003). Within this region were 6 potential StzA binding sites, with one being *REALALE* (-274 upstream of the ATG translation start site of the *prnB* gene; Figure 5.9). Putative overlapping StzA/AreA binding sites were also found in the *A. oryzae prnB* promoter at -186 (5'-AGATAAGGCA).

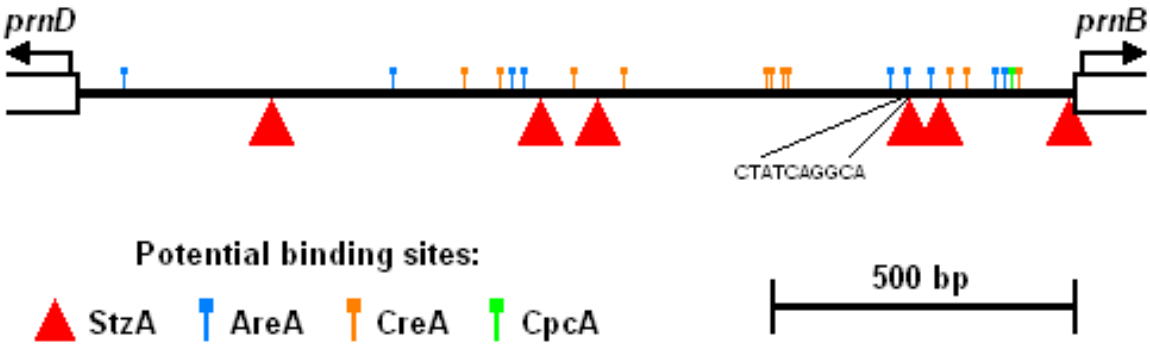


Figure 5.9. Positions of potential StzA binding sites in the *prnD–prnB* intergenic *cis*-acting regulatory region of the proline utilisation cluster in *A. nidulans*. All potential binding sites for the TFs StzA (red triangles below the line), AreA, CreA and CpcA (coloured markers above the line) are shown within this 1663 bp promoter region. Note the presence of the *REALALE* sequence (5'-CTATCAGGCA), consisting of putative overlapping StzA/AreA binding sites, at -274 upstream of the ATG translation start site of the *prnB* gene.

The *REALALE* sequence was found in the promoters of 5 genes, in addition to *stzA*, encoding zinc finger TF proteins (AN5924, AN1893, AN4013, AN8902 and AN8939). *REALALE* was also found within the promoter of a gene encoding a vacuolar ATPase subunit (AN5606) and the promoter of a gene involved in double-strand break repair *via* non-homologous end-joining (AN6705).

5.2.7c Identification of potential StzA target genes

The *A. nidulans stzA* gene plays a key role in the regulation of a wide variety of abiotic stress responses, including osmotic and DNA damage responses, cation transport and pH regulation (O'Neil *et al*., 2002; O'Mahony *et al*., 2002; Spielvogel *et al*., 2008; Findon *et al*., 2010). Furthermore, the *stzA* promoter analysis of the present study reveals the potential for the TFs involved in nitrogen metabolism (AreA) and amino acid biosynthesis (CpcA) to interact with *stzA* (Chilton *et al*., 2008). To identify possible genes that StzA targets, promoters of genes involved in these processes in *A. nidulans* were analysed for potential StzA binding sites. By chance alone, it would be expected that there be 3.91 StzA binding site motifs (AGGCA or its reverse complement TGCCT) in each 2 kb promoter analysed (1 site in each 512 bp). The presence of more sites than would be expected by chance within promoters may indicate a regulatory role for StzA. Physiologically relevant binding sites are often conserved in position between orthologues and/or arranged in tandem or multiple organisation (Ravagnani *et al.,* 1997). Comparison of promoter sequences from different species can aid identification of regulatory sequences because such sequences are typically more conserved than non-functional DNA (Wittkopp, 2010). Orthologues of *A. nidulans* genes were identified in *A. fumigatus* (or, if absent, the top BLAST match for an *Aspergillus* species) and were subjected to promoter analysis to assess if any binding sites present were positionally conserved among species. Where appropriate, additional promoters of orthologous genes in *Aspergillus* species were examined to assess the extent of the conservation.

5.2.7ci Osmotic stress response genes

A detailed study of the genes involved in the osmotic stress response in *A. nidulans* has been published by Miskei *et al*. (2009). These genes were analysed for potential StzA binding sites (Table 5.17). The promoter of the *sskB* gene (AN1180), which encodes a mitogen activated protein kinase kinase kinase (MAPKKK) of the HOG signaling pathway, contains sequences consisting of 2 StzA binding site motifs in similar positions in *A. nidulans* (5'-TGCCTGCCT at -1339) and *A. fumigatus* (5'-TGCCTGAGGCA at -1150). The gene that encodes the TF MsnA (AN1652), which also functions in the HOG pathway, contains a potential StzA binding site in its promoter that is conserved in position in 5 *Aspergillus* species: *A. nidulans* (-571), *A. fumigatus* (-822), *N. fischeri* (-818), *A. clavatus* (-793) and *A. terreus* (-708). This potential StzA binding site is part of the sequence 5'-TAGCACAGGCAGATC conserved among all 5 *msnA* promoters.

Considering that an average of approximately 4 StzA binding sites are expected to occur in each promoter by chance alone, some StzA binding site motifs appear to be heavily over-represented in many *A. nidulans* promoters of genes involved in the osmotic stress response. Notable examples include the promoters of genes encoding two histidine kinases: *tcsA* (AN5296) and *phkB* (AN3101), both containing 9 potential StzA binding sites. The promoter of *enaC* (AN7664), a plasma membrane-located Na^+ P-type ATPase, contains 7 potential StzA binding sites.

Table 5.17. Numbers of potential StzA binding sites in the promoters of *A. nidulans* genes involved in the osmotic stress response. The genes together with their proposed functions were obtained from Miskei *et al*. (2009).

Function	Accession number	No. of potential StzA binding sites
Sho1p-like stress sensing and signalling branch	AN7698 (*shoA*)	3
	AN7487 (*modA*)	4
	AN8836	0
	AN2067	2
	AN2269 (*steC*)	5
Histidine kinase	AN1800 (*tcsB*)	4
	AN3101 (*phkB*)	9
	AN3102	6
	AN5296 (*tcsA*)	9
	AN4479 (*nikA*)	4
Histidine-containing phototransfer intermediate	AN2005 (*ypdA*)	3
Response regulator	AN7697 (*sskA*)	4
	AN3688 (*srrA*)	5
MAPKKK	AN1180 (*sskB*)	5
MAPKK	AN0931 (*pbsB*)	5
MAPK	AN1017 (*hogA*/*sakA*)	5
	AN4668 (*mpkC*)	5
Regulation of the HOG pathway	AN6982 (*ptpA*)	3
	AN4896	4
	AN6892	6
	AN1358	3
Nuclear importin or exportin	AN5482 (*ranA*)	4
	AN6006	4
	AN1401 (*kapK*)	5
HogA / SakA targets	AN4483	5
	AN6505 (*rcoA*)	4
	AN2911 (*atfA*)	1
	AN1652 (*msnA*)	6
	AN4581	4
	AN4493 (*rpdA*)	6
AtfA targets	AN7664 (*enaC*)	7
	AN1628	4
	AN6642	3
	AN2911 (*atfA*)	1
	AN4896	4
	AN0351 (*gfdA*)	2
	AN6792 (*gfdB*)	2
	AN9339 (*catB*)	4
	AN2846	7
SsrA targets	AN0170 (*thiO*/*trxA*)	7
	AN3581 (*trxB*/*trxR*)	6
	AN7664 (*enaC*)	7
	AN1628	4
	AN6642	3
	AN4716	3
	AN9339 (*catB*)	4
	AN2846	7
Sgd1p target	AN0351 (*gfdA*)	2
Smp1p target	AN9168	0
MsnA targets	AN0351 (*gfdA*)	2
	AN9034	4
	AN1216 (*gppA*)	6
	AN9339 (*catB*)	4
RcoA targets	AN0351 (*gfdA*)	2
	AN9034	4
	AN9339 (*catB*)	4
	AN7664 (*enaC*)	7
	AN1628	4
	AN6642	3

5.2.7cii DNA repair genes

A comprehensive list of genes involved in DNA repair in *A. nidulans* has been published by Goldman and Kafer (2004). These genes were analysed for potential StzA binding sites (Table 5.18). It was apparent that the promoter of the AN6705 gene, which contains the *REALALE* sequence (at -846; Section 5.2.7b) also contains the potential double StzA binding site sequence 5'-TGCCTGAGGCA (at -316). Multiple StzA binding site motifs are present in conserved positions in the promoters of *ku80* gene orthologues (AN4552 in *A. nidulans*) in a region approximately -60 to -150 bp upstream of the ATG translation start sites (Figure 5.10). Furthermore, StzA binding site motifs appear to be over-represented in many *ku80* promoters, especially in *Aspergillus* species (Table 5.19).

Table 5.18. Numbers of potential StzA binding sites in promoters of *A. nidulans* genes involved in DNA repair. The genes together with their proposed functions were obtained from Goldman and Kafer (2004).

Function	Accession number	No. of potential StzA binding sites	Function	Accession number	No. of potential StzA binding sites
Photoreactivation	AN0387	3	Non-homologous end-joining	AN7753	3
Base excision repair (BER)	AN7435	7		AN4552	7
	AN3766	4		AN0097	3
	AN5479	6	Sanitization of nucleotide pools	AN0271	5
	AN6682	4	DNA polymerases and replication factors	AN1231	6
	AN7653	1		AN0040	5
Other BER factors	AN4736	4		AN7325	1
	AN0097	3		AN3067	3
	AN3129	6		AN0415	2
Direct reversal of DNA damage	AN4587	5		AN6303 (*uvsF*)	6
Repair of DNA protein crosslinks	AN5903	5		AN4789 (*uvsI*)	5
Chromatin structure	AN3468	5		AN2163	8
Mismatch excision repair (MMR)	AN5006	3		AN3909	3
	AN3749	4		AN4678	2
	AN1708	5		AN5620	3
	AN6316	4		AN2739	6
	AN0126	1		AN3373	9
	AN4365	4		AN1231	6
Nucleotide excision repair (NER)	AN6186	3		AN6114	3
	AN2304	7		AN5694	4
	AN3784	2	Editing and processing nucleases	AN2764	4
	AN7423	4		AN3035	6
	AN0582	5		AN8259	6
	AN8201	6		AN3118	6
	AN9436	4	Rad6 pathway	AN5344 (*uvsJ*)	3
	AN5569	4		AN7309	6
	AN4283	8		AN3644	2
	AN0138	6		AN8702	3
	AN7015	2	Genes defective in diseases associated with sensitivity to DNA-damaging agents	AN2087 (*musN*)	2
	AN8285	4		AN0038	3
	AN2211	6	Other identified genes with a suspected DNA repair function	AN7589	10
	AN1668	6		AN3958	6
	AN5216	4		AN5751	7
	AN4331	6	Other conserved DNA damage response/cell cycle check point genes	AN6975 (*uvsB*)	3
	AN8713	9		AN5165 (*uvsD*)	2
	AN6069	6		AN3620	1
	AN4272	5		AN6616 (*nuvG*)	3
NER-related	AN5235	1		AN5494	3
	AN7103	4		AN4279	4
	AN0111	7		AN6896	3
	AN0596	3		AN0044	6
	AN1528	2		AN1971	3
	AN0604	2		AN8876	4
Homologous recombination	AN1237 (*uvsC*)	5		AN2969	3
	AN9092	6		AN6300	0
	AN4407 (*radC*)	2		AN8064	5
	AN5527	3		AN6517	5
	AN0855	4			
	AN1051	1			
	AN3619	3			
	AN0556 (*mreA*)	5			
	AN3372 (*scaA*)	9			

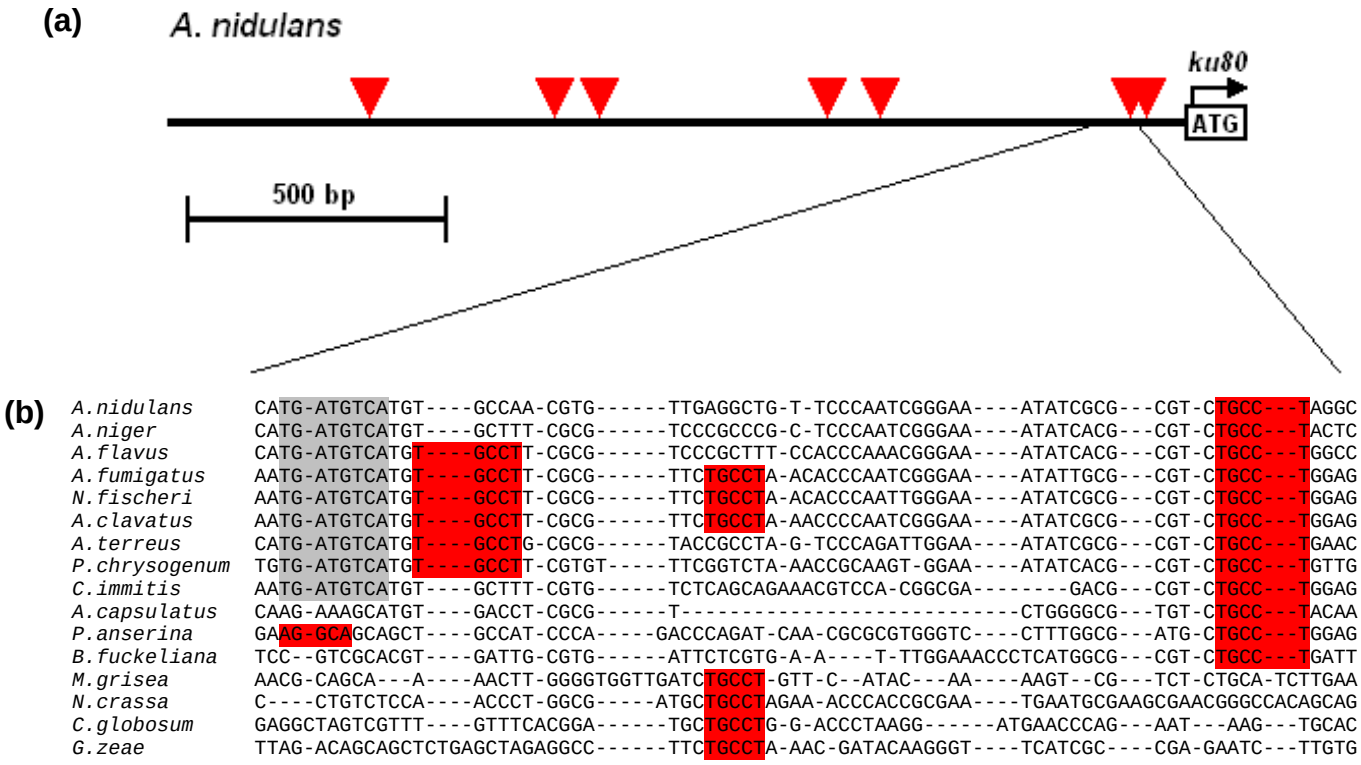


Figure 5.10. Conservation of potential StzA binding sites in the promoters of *ku80* orthologues. (a) Potential StzA binding sites in the promoter of the *A. nidulans ku80* gene are shown as red triangles. (b) A ClustalW alignment reveals that the *ku80* promoters of many filamentous ascomycetes contain StzA binding site motifs (highlighted in red) in one or more of three conserved positions approximately -60 to -150 bp preceding the ATG translation start sites. Within this region, the reverse complement of a consensus sequence of TGAC(G/A)TCA (highlighted in grey) identified in the promoters of SskA-, HogA- and AtfA-dependent fludioxonil up-regulated genes by Hagiwara *et al*. (2009) was found to be present and conserved in position in *Aspergillus ku80* promoters.

Table 5.19. Numbers of potential StzA binding sites within the promoters of *ku80* orthologues.

Species	Accession number	No. of potential StzA binding sites in *ku80* promoters
Aspergillus nidulans	AN4552	7
Aspergillus niger	ANI_1_1864064	6
Aspergillus flavus	AFLA_133460	11
Aspergillus fumigatus	AFUA_2G02620	7
Neosartorya fischeri	NFIA_035020	7
Aspergillus clavatus	ACLA_092170	7
Aspergillus terreus	ATEG_06919	5
Penicillium chrysogenum	Pc18g01820	4
Coccidioides immitis	CIMG_01804.3	3
Ajellomyces capsulatus	EER44165	5
Uncinocarpus reesii	UREG_05499	8
Phaeosphaeria nodorum	SNOG_11900	8
Pyrenophora tritici-repentis	PTRG_00307	5
Podospora anserina	PODANSg5328	5
Botryotinia fuckeliana	BC1G_10199	4
Magnaporthe grisea	MGG_10157	6
Neurospora crassa	NCU00077	4
Chaetomium globosum	CHGG_05790	5
Gibberella zeae	FG06721.1	6

5.3.7ciii pH regulation and ion homeostasis genes

Key genes involved in pH regulation have been summarised by Peñalva *et al*. (2008). Numbers of potential StzA binding sites in the promoters of genes involved in pH regulation are shown in Table 5.20. PacC is the main pH-responsive TF in *A. nidulans* (Tilburn *et al*., 1995). The *pacC* promoter in *A. nidulans* (AN2855) contains 8 potential StzA binding sites, which is double the number expected by chance alone. The *pacC* promoter also contains 12 potential PacC binding sites (5'-GCCARG) and 2 of these overlap with potential StzA sites at -388 (GCCAAGGCA) and -1305 (GCCAGGCA). Hence, interactions of StzA and PacC may function in the regulation of *pacC* by binding site exclusion.

palA is one of six genes participating in pH signal tranduction in *A. nidulans* (Negrete-Urtasun *et al*., 1997). The *palA* promoter in *A. nidul*ans contains a potential double StzA binding sequence at -112 (TGCCTAAGGCA), which is conserved in similar positions in the promoters of *A. fumigatus* (-160), *N. fischeri* (-118), *A. clavatus* (-102) and *A. niger* (-98; Figure 5.11).

Table 5.20. Numbers of potential StzA binding sites within the promoters of *A. nidulans* genes involved in pH regulation. Genes involved in pH regulation were obtained from Peñalva *et al*. (2008).

Gene	Accession number	No. of potential StzA binding sites
pacC	AN2855	8
palA	AN4351	6
palB	AN0256	4
palC	AN7560	3
palF	AN1844	7
palH	AN6886	7
palI	AN4853	2
vps32	AN7310	4

```
A.nidulans      GTAAATACAG-ATAGC-TGCCTAAGGCACT-ACCACGTCATTTGCCCGCCT  -112
A.fumigatus     GAAGGAAGGACATGGAATGCCTCAGGCAGAGAGGACGTCATCC-GTTGCCA  -160
N.fischeri      GACGGAAGGACGTGGAATGCCTCAGGCAGA-AGGACGTCATCC-GTTGCCA  -118
A.clavatus      AATATGAGGGTTTGGC-TGCCTCAGGCATCGAGACCGTCATCCCGTTGCCA  -102
A.niger         GAGTGTAACGTACTTTCTGCCTCAGGCACCCACAACGTCATGC-GTTGCCC   -98
                      *          ***** *****   *   ******      ***
```

Figure 5.11. ClustalW aligment showing double conserved potential StzA binding sites in the promoters of *Aspergillus palA* genes. A sequence (5'-TGCCTMAGGCA) containing double StzA binding site motifs is conserved in sequence and position in 5 *Aspergillus* species and a very similar sequence is present in *A. terreus* (5'GGCCTGAGGCA at -111).

Transporters involved in ion homeostasis have been identified for *A. nidulans* based upon homologies when compared to their yeast counterparts (Benčina *et al*., 2009). These promoters were analysed for potential StzA binding sites (Table 5.21). It was evident that the genes AN4920 (*pmcB*), AN4131, AN5743 and AN7510 have at least double the numbers of expected StzA binding sites in their promoters, which may indicate functional significance.

Table 5.21. Numbers of potential StzA binding sites in promoters of *A. nidulans* genes involved in ion homeostasis. The genes identified were based upon homologies when compared to their yeast counterparts (Benčina *et al*., 2009). The names of ion transporters that have been characterised experimentally are shown in brackets.

Type of ion transporter	Accession number	No. of potential StzA binding sites
K^+, Na^+/H^+ exchanger	AN1290	6
	AN1920	6
	AN2288	6
	AN4131	9
	AN5035	6
	AN5187	4
	AN7250 (*nha1*)	3
Ca^{2+}/H^+ exchanger	AN0471 (*vcxA*)	2
	AN4266	3
	AN5821	2
	AN6986	3
	AN7173	4
	AN7510	8
P-type H^+ and Ca^{2+} ATPase	AN0318	6
	AN1189 (*pmcA*)	3
	AN1628 (*enaB*)	4
	AN2827	2
	AN4859 (*pmaA*)	5
	AN4920 (*pmcB*)	10
	AN5088	3
	AN5743	9
	AN6642 (*enaA*)	3
	AN7464 (*pmrA*)	1
	AN7664 (*enaC*)	7
	AN8399	4

5.2.7civ Nitrogen metabolism genes

Principal *A. nidulans* genes involved in nitrogen metabolism, including those involved in nitrate assimilation and cross pathway control of amino acid biosynthesis, were analysed for the presence of StzA binding site motifs (Table 5.22).

Table 5.22. Numbers of potential StzA binding sites in promoters of *A. nidulans* genes involved in nitrogen metabolism. Genes were obtained from McCahill *et al*. (2002); Krappmann and Braus (2005).

Pathway	Gene	Accession number	No. of potential StzA binding sites
Nitrate assimilation	*areA*	AN8667	6
	areB	AN6221	5
	niiA	AN1007	4
	nirA (niiB)	AN0098	3
	niaD	AN1006	4
	nrtA (*crnA*)	AN1008	9
	nrtB	AN0399	6
	nitA	AN8647	4
	meaB	AN4900	4
Cross pathway control of amino acid biosynthesis	*cpcA*	AN3675	7
	cpcB	AN4163	5
	cpcC	AN2246	2

.

Subsequently, potential StzA binding sites were identified in the *cpcA* and *cpcB* promoters of both *A. nidulans* and *A. fumigatus* (Figure 5.12). Many StzA binding sites demonstrated conservation of position relative to the ATG translation start site of each gene, with 2 sites forming a doublet of 5'-TGCCTN$_{(12-18)}$TGCCT in *cpcA* promoters. The results of analysing the *cpcA* promoters from a wider range of fungi are summarised in Table 5.23, where this doublet structure was observed in 11 of the promoters analysed, and the degree of conservation is shown in Figure 5.13. The presence of this sequence tends to correlate with an over-representation of StzA binding site motifs in many *cpcA* promoters (Figure 5.14). A potential CpcA binding site was also found in each *cpcA* promoter, indicating that transcriptional auto-regulation contributes to CpcA expression (Krappmann and Braus, 2005). The results obtained seem especially significant considering that it has been shown that potential CpcA binding sites are present in 26 out of 29 *stzA* promoters in a conserved position -55 to -169 of the 5' ATG translation starts (Figure 5.8b). Furthermore, the *A. nidulans* and *A. fumigatus cpcB* promoters both contain 5 potential StzA binding sites and several of these

show conservation of position (Figure 5.12), indicating a further possible role for StzA in the regulation of *cpcB*. *cpcA* and *cpcB* play antagonistic roles in cross pathway control and sexual development (McCahill *et al*., 2002).

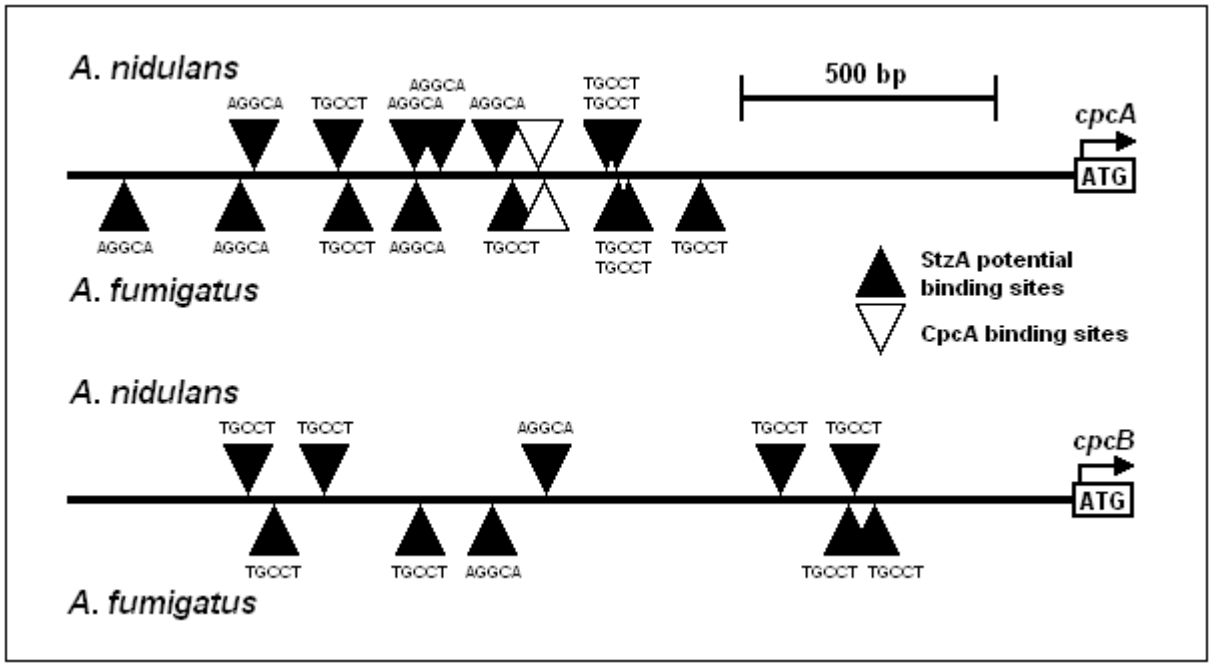


Figure 5.12. Conservation of potential StzA binding sites in *cpcA* and *cpcB* promoters of *A. nidulans* and *A. fumigatus*. Note that within the *cpcA* gene, two StzA sites form a doublet of 5'-TGCCTN$_{(12\text{-}19)}$TGCCT, conserved in the promoters of 11 *cpcA* orthologues.

Table 5.23. Summary of potential StzA binding sites in *cpcA* orthologues and CpcA binding sites in *stzA* orthologues. ([a]) represents a double conserved AGGCA binding site motif. ([b]) represents a double conserved TGCCT-AGGCA binding site motif. N/D = Not Determined.

Species	*cpcA* accession number	Number of potential StzA binding sites in *cpcA* promoters	Position of potential double StzA binding sites in *cpcA* promoters (bp between sites)	Position of potential CpcA sites in *stzA* promoters
Aspergillus nidulans	XM_656187	7	-911 (18)	-67
Aspergillus fumigatus	XM_746491	8	-889 (13)	-107
Aspergillus niger	X99215	6	-926 (13)	-96
Neosartorya fischeri	XM_001266789	7	-889 (13)	-107
Aspergillus flavus	XM_002383111	7	-870 (13)	-1152, -1586
Aspergillus oryzae	XM_001816865	7	-870 (13)	-1152, -1586
Aspergillus clavatus	XM_001272098	7	-913 (15)	-91
Aspergillus terreus	XM_001212309	5	-885 (19)	-1362
Ajellomyces capsulatus	XM_001538459	8	-375 (12)	-139
Uncinocarpus reesii	XM_002585089	2	Not seen	-87
Talaromyces emersonii	-	N/D	N/D	-87, -151
Penicillium chrysogenum	-	N/D	N/D	-89
Arthroderma benhamiae	ARB_03684	4	Not seen	-90
Arthroderma gypseum	MGYG_04606	3	Not seen	-91
Trichophyton rubrum	TERG_08022	4	Not seen	-90
Paracoccidioides brasiliensis	PAAG_03094	2	Not seen	-140
Coccidioides immitis	XM_001243155	1	Not seen	-79
Phaeosphaeria nodorum	SNOG_07085.1	3	Not seen	-55
Gibberella zeae	XM_389462	6	-404 (7) [a]	-144
Trichoderma reesei	Tre37844	4	-859 (15)	-169, -305
Chaetomium globosum	XM_001224922	4	Not seen	-123
Neurospora crassa	XM_952572	4	-785 (38)	-119
Magnaporthe grisea	XM_368642	11	-231 (16)	-119
Podospora anserina	-	N/D	N/D	-130
Nectria haematococca	-	N/D	N/D	-146
Sordaria macrospora	-	N/D	N/D	-119
Verticillium albo-atrum	VDBG_08221	9	Not seen	-127
Botryotinia fuckeliana	XM_001556388	3	Not seen	-93
Pyrenophora tritici-repentis	PTRG_00426	10	-891 (20) [b]	-58

```
A.fumigatus     AGGGGTTTTAAGAAGTGCCTGATTTCCC------GTCGCTGCCTGGCGCTTTTCTTTTC  -889
N.fischeri      GAGGGTTTTAAGAAGTGCCTGATTTCCC------GTCGCTGCCTGGCGCTTTTCCTTTC  -889
A.clavatus      GAGGGTTTTAAGAAGTGCCTGATTTCCCCC----GTCGCTGCCTGGCGCTTCTTCTCTC  -913
A.niger         CAAACAGTTAAGAAGTGCCTGGCTTTCC------GTCGCTGCCTGGCGCTTCTTTTCTC  -926
A.terreus       CACCCGCTTAAGAAGTGCCTGATTTCCCCCTCCCGTCGTTGCCTGGCGTTTTTCTTTTC  -885
A.oryzae        CTGCG-TTTAAGAAGTGCCTGATTTCCC------ATCGTTGCCTGGGCTTCTTTTTCCT  -870
A.flavus        CTGCG-TTTAAGAAGTGCCTGATTTCCC------ATCGTTGCCTGGGCTTCTTTTTCCT  -870
A.capsulatus    CCGCGCTTTAATAGATGCCTGATT-CCC------ACTTTTGCCTTTTCTCCTTCCTCAT  -375
A.nidulans      AAGACTGTTAAAAGGTGCCTTATCCCCCGG-CACGTCGCTGCCTGGCTTTCCTCCCTGA  -911
T.reesei        GTCCGCC-CAGCGGGTGCCTAACA-GCCTG---ACGCGCTGCCTGGC-GCCGGTGCGTG  -859
M.grisea        TTCCGCCATCGGAAGTGCCTCTCGCAACCA---CCCTTCTGCCTCGCCATCAACGCAGA  -231
                          *****       *          *****
```

Figure 5.13. ClustalW alignment of potential double conserved StzA binding sites in *cpcA* promoters. Red highlighted regions indicate potential StzA binding sites and (*) indicates conservation of nucleotides across the selected species. Numbers indicate the position of the TGCCT doublet motif upstream of the ATG translation start sites from the nucleotide highlighted in bold (**T**).

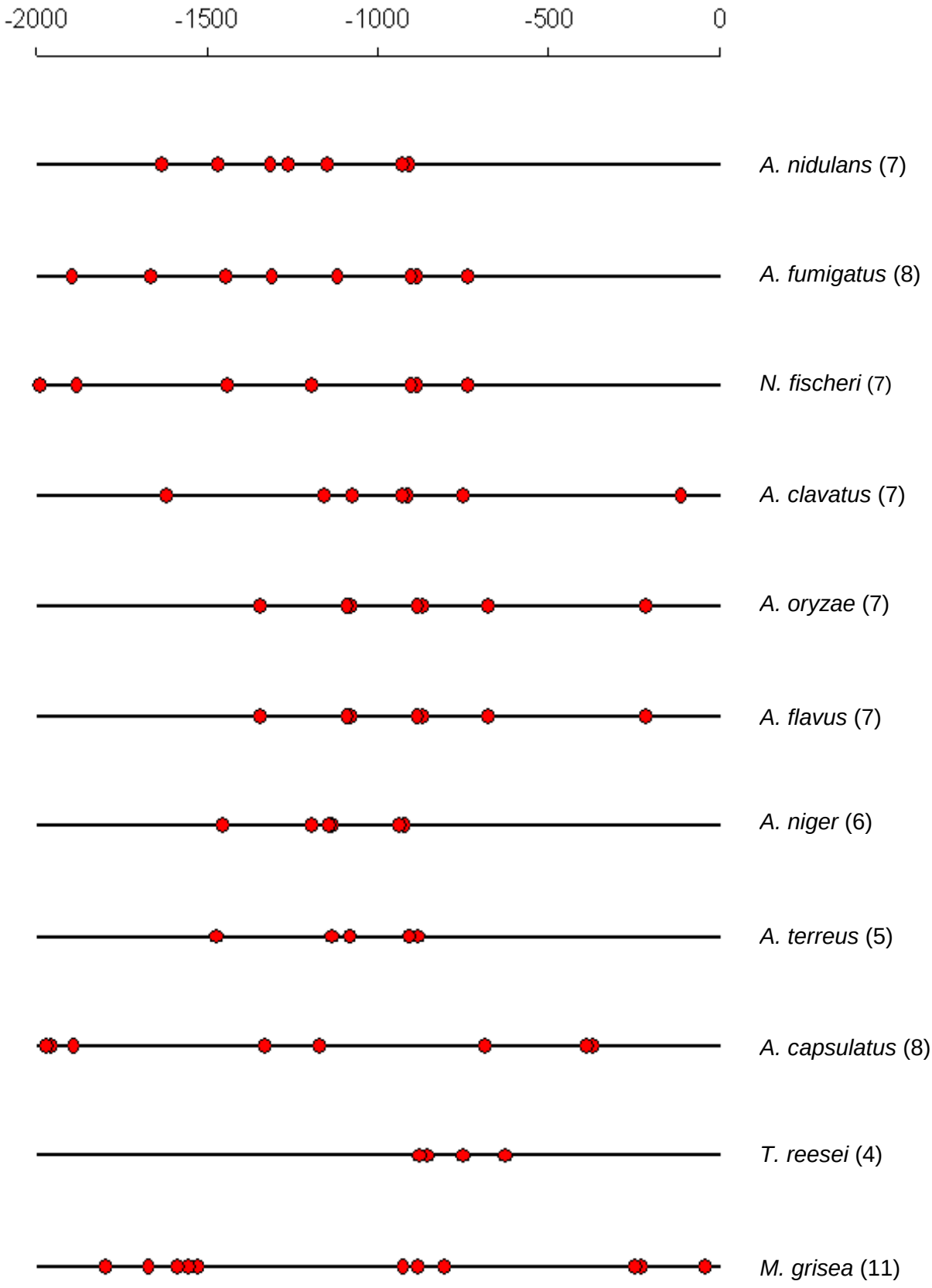


Figure 5.14. Positions of potential StzA binding sites in *cpcA* promoters. Red spheres represent StzA binding site motifs, and numbers along the top show their positions upstream of the ATG translation start sites. Since a number of StzA binding sites are in close proximity, the total number of sites for each promoter is shown in brackets after the species name.

5.2.7d Statistical analysis of StzA binding site motifs within *A. nidulans* promoters

Chi-square analysis was used to assess whether the frequency distribution of StzA binding site motifs was significantly different in any groups of *A. nidulans* promoters from genes assigned to functionally related groups when compared to a control group of promoters. This approach was used by Coutinho *et al*. (2009) to assess if any groups of promoters from genes assigned to functionally related groups were enriched for specific TF binding site motifs. Groups of genes used in the analysis were those involved in osmotic stress, DNA repair, cation homeostasis and those containing the *REALALE* sequence within their promoters (Tables 5.16–5.18, 5.21 and Figure 5.15). Genes from the control group were chosen at random from the *A. nidulans* genome by selecting every 125th gene based on accession numbers (Appendix 8). In some cases, smaller categories were combined for the chi-square analysis so that the combined expected values were 5 or more to satisfy the criteria of the test. However, it was noted that the results obtained were similar even when categories were not combined, which demonstrates the robustness of the analysis. The histogram in Figure 5.15 shows the frequency with which different numbers of StzA binding site motifs occur in the promoters of genes assigned to functionally related groups in *A. nidulans*. A summary of the results obtained for the chi-square analysis is shown in Table 5.24, which is detailed further in Appendix 10.

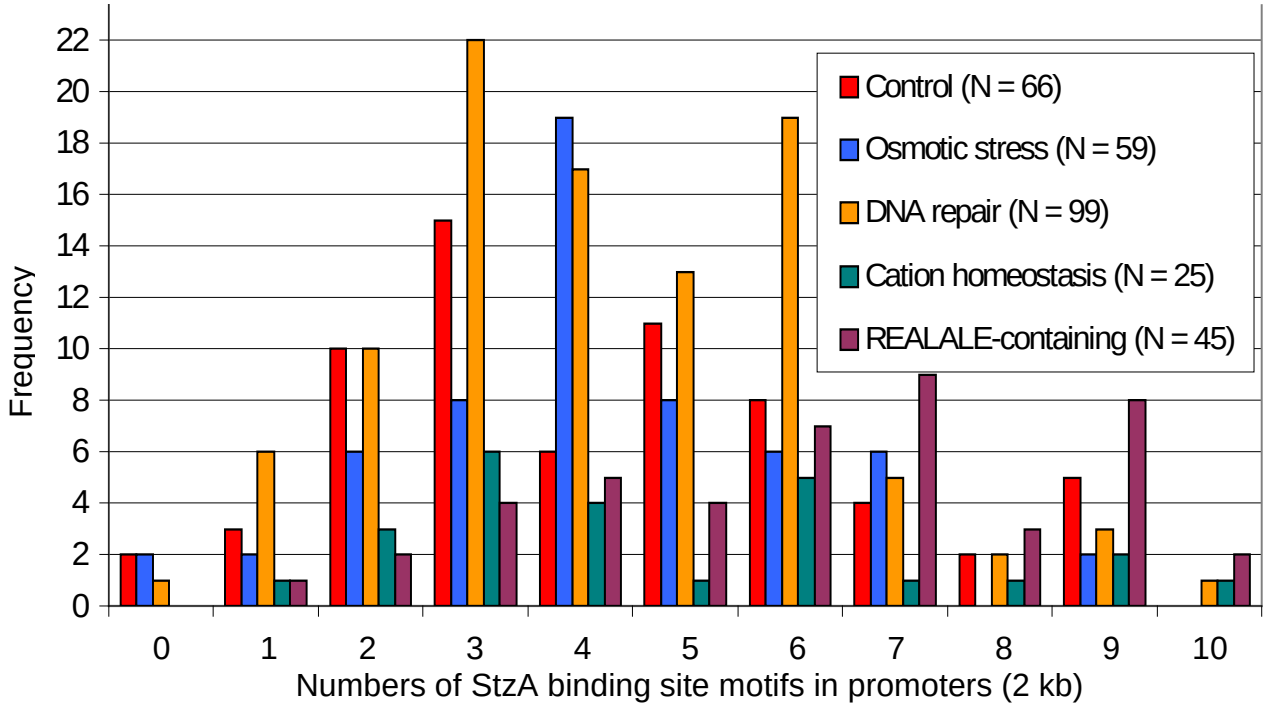


Figure 5.15. The frequency with which different numbers of StzA binding site motifs occur in the promoters (2 kb) of genes assigned to functionally related groups in *A. nidulans*.

Table 5.24. Chi-square values for the frequency distributions of groups of genes assigned to functionally related groups when compared to a control group of genes. The full analysis is shown in Appendix 10.

	Functional group				
	Control	**Osmotic stress**	**DNA repair**	**Cation homeostasis**	***REALALE*-containing**
N	66	59	99	25	45
Mean	4.30	4.24	4.31	4.76	6.18
Chi-square value (X^2)	-	42.98	15.74	1.92	31.22
Critical value (P = 0.05)	-	12.59	14.07	7.81	9.49

Statistically, it would be expected that a random sequence for an StzA binding motif occurs once in every 512 bp (4^5 / 2) in promoters, assuming equal occurrences of the nucleotides A, T, C and G in fungal promoters. This is equates to 3.91 occurrences in every 2 kb promoter. The actual mean occurrence of StzA binding motifs in promoters of the control group of genes is 4.30 per 2 kb (244 StzA binding motifs in 66 promoters). This is therefore slightly higher compared to chance alone and may be due to some genes in the control group containing functional StzA binding sites. For promoters of genes involved in the osmotic stress response in *A. nidulans*, the calculated value for χ^2 (42.98, d.f. = 6, P<0.05) is above the critical value (12.59), and therefore the frequency distribution of StzA binding sites in this group of genes differs significantly from the expected ratio derived from the control group. This significance can be attributed to the differences in the frequencies of promoters with 4 StzA binding site motifs between these two groups of genes. For promoters of genes involved in the DNA damage response, the calculated value for χ^2 (15.74, d.f. = 7, P<0.05) is above the critical value (14.07). This is indicative of statistical significance for the frequency distribution of StzA binding site motifs among genes involved in the DNA repair response compared to the control group. For the group of genes involved in cation homeostasis, the calculated value for χ^2 (1.92, d.f. = 3, P>0.05) is below the critical value (7.81). This demonstrates that there is no statistical significance between the frequency distribution of StzA binding site motifs between genes involved in cation homestasis and the control group when considering the group of cation transporters as a whole. For *A. nidulans* gene promoters containing the *REALALE* sequence, the calculated value for χ^2 (31.22, d.f. = 4, P<0.05) is above the critical value (9.49) and therefore the frequency distribution of StzA binding site motifs for this group differs significantly from the control group. This significance is largely due to the greatly increased frequency of promoters with 6 or more StzA binding motifs in the *REALALE*-containing group of genes compared to the control group. This is evident by a frequency distribution of StzA binding site motifs that is skewed to the right

for the *REALALE*-containing group (Figure 5.15). The mean occurrence of 6.18 StzA binding site motifs per 2 kb promoter in *REALALE*-containing promoters is also substantially more than 4.30 for the control and 4.24 to 4.76 for other groups of genes (Table 5.24).

5.3 Discussion

5.3.1 Taxonomic distribution of StzA

Database searches have revealed that StzA and its orthologues appear to be restricted to filamentous Ascomycota (Tables 5.1–5.3). The Ascomycota can be divided into three subphyla: the Taphrinomycotina, Saccharomycotina and Pezizomycotina (Lutzoni *et al*., 2004). It seems that StzA was acquired after the divergence of the Pezizomycotina subphylum (last to diverge) since no genomes from the Saccharomycotina and Taphrinomycotina subphyla were found to contain StzA orthologues. The first class to diverge within the Pezizomycotina was the Pezizomycetes, which contains a high proportion of mycorrhizal fungi (Berbee, 2001). No resources were available to search representative species from this class of fungi. The next class to diverge (the Leotiomycetes) contains *B. cinerea*, which possesses an StzA orthologue. All other classes of Pezizomycotina for which data is available also possess StzA orthologues. This indicates the *stzA* gene was acquired either between the Saccharomycotina–Pezizomycotina divide or between the Pezizomycete divide from the other Pezizomycotina. Data from a Pezizomycete genome project would help answer this question, as well as indicating whether StzA could play a role in a mycorrhizal environment.

The Pezizomycotina subphylum contains highly diverse fungal species, including numerous species that impact on humans and plants in both beneficial and detrimental ways (Tree of Life Web Project, 2006; http://tolweb.org/Pezizomycotina/29296). Within the Pezizomycotina are 4 classes: the Dothideomycetes, Eurotiomycetes, Sordariomycetes and Leotiomycetes. Wang *et al*. (2009) unambiguously place the Dothidiomycetes sister to the Eurotiomycetes and the Sordariomycetes sister to the Leotiomycetes by using more organisms from the Dothideomycetes class to counteract the conflicting results of Fitzpatrick *et al*. (2006) and Robbertse *et al*. (2006). Species containing StzA orthologues fall into 6 orders within these 4 classes (Table 5.3). Since StzA is present in many animal and plant pathogens, and a pathogenic existence poses particularly stressful environments (Berbee, 2001), StzA could play a role in enabling fungi to adapt to stresses encountered in these environments. Baker *et al*. (2004) have shown that fungi in mixed communities in a hostile acid mine environment (pH 1, 30–50 ºC, containing 0.3 M metal salts, mostly iron, but with significant amounts of arsenic and copper), all appeared to be from the Pezizomycotina subphylum, in particular from the Eurotiomycetes and Dothidiomycetes classes.

5.3.2 *stzA* intron annotation errors

Introns are often incorrectly predicted using automated annotations alone. Rep e*t al.* (2006) estimated that 1 in 10 genes in fungal genomes are likely to be incorrectly annotated due to the limitations of gene prediction programs currently available. Since the majority of protein-coding genes in sequenced genomes have not been experimentally characterised, manual annotations of genes are a potentially more accurate method for correctly identifying introns than automated intron prediction methods. Predictions of *stzA* genes and their corresponding proteins solely by automated computer analysis is indicated by the fact that these sequences that originate from genome sequencing projects are represented by the accession numbers XM_ and XP_, which indicate predictions for mRNA and protein sequences, respectively (Peri *et al*., 2001). These sequences, obtained from BLAST searches (Tables 5.1 and 5.2), were examined for potential misannotations. Using DNA and protein alignments, the known conservation of intron positions and information regarding fungal intron structure, it was possible to provide manual *stzA* annotations that were potentially more accurate (Figure 5.2). It became apparent that most of the genes had been misannotated. A total of 21 out of 86 introns were found to be misannotated at NCBI (Table 5.4) and 28 out of 69 introns misannotated at the Broad Institute (Table 5.6). A further predicted intron appeared to be misidentified in the C-terminus of *G. zeae stzA* at the Broad Institute. Additionally, at the Broad Institute, the N-terminus was missing from the *stzA* genes of *C. immitis* (923 bp) and *B. cinerea* (96 bp). Following manual annotation, 16 (out of 29) and 19 (out of 22) *stzA* genes and their corresponding proteins differed in sequence at NCBI and the Broad Institute, respectively (Table 5.5).

Incorrect annotation of *stzA* orthologues by computer analysis resulted mainly from missed introns and under or overestimation of intron sizes due to incorrect prediction of 3' or 5' splice sites (Table 5.5). Proteome comparisons among the genomes of *A. nidulans*, *A*. *fumigatus* and *A. oryzae* have revealed major differences in gene model annotation. The majority of identified orthologue groups (~80%) contained members differing in length and/or numbers of exons (Wortman *et al*., 2009). Identified gene annotation problems included the merging of neighbouring loci, missed exon calls and incorrect 5' exons. Inconsistencies in gene annotation were attributed to the diverse annotation processes employed at different sequencing centres and the lengthy timeframe over which the different genomes were sequenced. Genome annotation methods used by sequencing centres associated with *Aspergillus* species have been summarised by Wortman *et al*. (2009) and Haas *et al*. (2011). Some sequencing groups predicted the functions of gene

products based on high stringency homology matches to experimentally characterised proteins, whereas other groups applied more lenient criteria. It has been observed that 30% of *A. nidulans* genes overlapping expressed sequence tag (EST) alignments showed some inconsistency with sequence data (Wortman *et al*., 2009). Many of the inconsistencies were due to the merging of neighbouring loci, as revealed by comparative analysis of *A. nidulans* genes with their orthologues in *A. fumigatus* and *A. oryzae*.

These findings are perhaps not surprising considering that even in the well-characterised yeast *S. cerevisiae*, 29.9% (26 out of 87) intron predictions were found to be incorrect (Davis *et al*., 2000b). A comparison of the *S. cerevisiae* genome with that of the filamentous ascomycete *Ashbya gossypi* identified a number of unrecognised introns/exons within the *S. cerevisiae* genome and revealed that many ORFs had been overestimated in size due to the incorrect prediction of 3' or 5' splice sites (Brachat *et al*., 2003). Furthermore, in *S. cerevisiae* several sequencing errors were exposed and STOP codons were found to be wrongly predicted that resulted in adjacent pairs of genes being incorrectly fused. Thus, despite the fact that *S. cerevisiae* is considered to be one of the most accurately sequenced and annotated genomes, a substantial number of gene annotation errors (95) were identified within its genome by Brachat *et al*. (2003). These errors were confirmed by comparisons with gene orthologues from *C. albicans*, *S. pombe* and *N. crassa* obtained by BLAST searches.

Comparative genomics approaches have demonstrated that there are considerable benefits to comparing closely related species to improve gene annotation (Davis *et al*., 2000b; Brachat *et al*., 2003; Wortman *et al*., 2009). Hence, the forthcoming availability of further completely sequenced genomes for filamentous fungi could be used to verify the correct manual annotation of *stzA* orthologues and may lead to further improvements in their annotation. Furthermore, verification of the correct manual annotation of introns for selected *stzA* orthologues could be achieved by using experimental techniques such as 5' rapid amplification of cDNA ends (5' RACE) and subsequent sequencing.

It has been noted that a number of proposed changes to *S. cerevisiae* genes that have been reported in the literature failed to be corrected on databases (Brachat *et al*., 2003). Likewise, since the Broad Institute database searches were checked and updated in October 2012, it appears that the proposed manual changes to *stzA* gene/StzA protein annotations published in Chilton *et al*. (2008) have yet to be incorporated into the databases, considering that 19 of the 22 *stzA*/StzA orthologues available remain incorrectly

annotated (Table 5.6). As the Broad Institute is responsible for producing and analysing the sequence data from fungal organisms that are part of the the Fungal Genome Initiative (FGI; http://www.broadinstitute.org/annotation/fungi/fgi/), it seems particularly important that this website is updated regularly with the manual annotations to fungal genes/proteins presented in published papers.

All introns, whether identified by computer analysis or annotated manually, belonged to the GU-AG canonical class that accounts for over 98% of fungal introns (Figure 5.3; Kupfer *et al*., 2004). Fifty-eight out of 86 introns (67.4%) exactly matched the fungal 5' splice site consensus sequence GURWGU and 59 introns (68.6%) matched the fungal 3' splice site consensus sequence YAG (Table 5.7; Kupfer *et al*., 2004). Approximately half of the introns (42 / 86) exactly matched the fungal branchpoint consensus sequence RCURAY (Kupfer *et al*., 2004). Intron size ranged from 45 to 86 nucleotides. These sizes are in broad agreement with the findings of Kupfer *et al*. (2004) that show *A. nidulans* and *N. crassa* genes having narrow intron length ranges with a dominant peak distribution between 50 and 70 nucleotides.

Due to the conserved N-termini and zinc finger regions within StzA proteins and the previously experimentally determined introns, it was relatively easy to assign introns into four positions (Figures 5.1 and 5.3). This does not, however, preclude the possibility of further introns in the C-termini. The introns showed the 5' bias of relatively intron poor organisms, such as fungi (Nielsen *et al.*, 2004; Jeffares *et al.*, 2006). The mechanisms and forces that underlie the evolution of introns are far from certain (reviewed in Roy and Gilbert, 2006), but it is possible to speculate about patterns of intron gain and loss among the *stzA* orthologues. Introns can be lost by simple DNA deletion or *via* a spliced RNA intermediate (Stajich and Dietrich, 2006). The most likely explanation for the absence of introns 2 and 3 in *A. nidulans* and *M. grisea,* respectively, is intron loss, and given the conservation of amino acids around these regions (Appendix 6), it seems likely to have occurred *via* an RNA intermediate. The same explanation may apply for the likely loss of intron 3 from the three members of the Onygenales Arthrodermataceae family (*A. benhamiae*, *A. gypseum* and *T. rubrum*). It is possible that the absence of intron 4 from members of all 4 classes is the result of separate losses from all four classes. However, it is more parsimonious to suggest that intron 4 is the result of an intron gain by the Eurotiomycetes and then intron losses to *A. nidulans* and *P. brasiliensis*. Intron gain is thought to most commonly occur by intron transposition, whereby an intron reverse splices into an mRNA transcript (reviewed in Roy and Irimia, 2009). The absence of intron 1 in the Sordariomycetes may be due to either an intron loss in this class or an intron gain by the Eurotiomycetes.

In summary, the manual annotations required for *stzA* orthologues seem to reflect the fact that annotation programs are lagging behind in their ability to accurately pinpoint coding regions compared to the vast increases in sequence data that have recently become available (Wortman *et al*., 2009). The gene structure annotations of *stzA* genes predicted by computer annotation should improve over time as new computational approaches emerge. However, gene predictions, whether relying on advanced computer annotation prediction methods alone or subject to additional manual annotation, must always be confirmed by "wet lab" experimentation. This is the only way to confirm the existence of a putative gene and determine the functionality of its protein product (reviewed in Sleator, 2010). It is also possible that more than one protein could be translated for StzA orthologues by the process of alternative splicing, in which exons are shuffled to generate novel proteins (reviewed in Schellenberg *et al*., 2008). Alternative splicing is thought to occur in all eukaryotes, including yeasts and fungi, despite their paucity of introns and short intron lengths (McGuire *et al*., 2008). Therefore, the existence of several versions of mRNA transcripts as a result of alternative splicing should also be considered when deducing protein sequences using databases of ESTs (Sims *et al*., 2004).

5.3.3 Conserved sequences within StzA proteins

StzA is a three-domain TF, consisting of an uncharacterised N-terminal domain, a central DNA-binding domain and an acid-rich putative C-terminal transcriptional activator domain (O'Neil *et al*., 2002). Although no structural functions for the N-terminal or C-terminal regions have been established for any StzA proteins using Conserved Domain Database (CDD) searches at NCBI, it has been shown that the zinc fingers share conserved motifs with a number of TFs. These include CreA from *A. nidulans*, a zinc finger protein from *S. cerevisiae*, a zinc finger AT-hook protein from *Drosophila melanogaster* and the human Wilm's tumour susceptibility protein WT-1 (O'Neil *et al*., 2002; Barham-Morris, 2006).

It is evident that the regions containing the zinc fingers and putative NLS are highly conserved among StzA proteins from filamentous ascomycetes (Figure 5.4). The secondary and tertiary structures of the triple C_2H_2 zinc finger binding domain is well established (reviewed in Knight and Shimeld, 2001). This domain in StzA has the structural motif $CN_{2/4}CN_{12}HN_3HN_5CN_4CN_{15}HN_4HN_8CN_4CN_{12}HN_4H$. Interestingly, the Sordariomycetes and the one representative Leotiomycete (*B. cinerea*) have 4 amino acid residues between the paired cysteines of the first zinc finger, whereas the Eurotiomycetes and Dothideomycetes have only 2

amino acid residues within this region. All zinc fingers conform to the $CN_{2-4}CN_{12}HN_{3-5}H$ consensus (Jacobs, 1992) with the exception that the middle zinc finger has 15 residues between the C and H residues. At least one sequencing error is apparent in the zinc finger region of *V. albo-atrum* since the second cysteine residue of the third zinc finger is replaced by leucine.

The most likely region of the NLS in StzA orthologues is shown highlighted in green in Figure 5.4. This sequence in Ace1 of *T. reesei* was predicted to be RRKKNATPEDVAPKKCR (Saloheimo *et al*., 2000). Comparison of this sequence (underlined) with orthologous sequences in the same region indicates that the bipartite NLS in StzA orthologues begins with the four basic amino acids sequence RRKK or RRRK and ends with a much less well-conserved sequence, which overlaps the first carbon atom of the C_2 of the first zinc finger. Note that the number of basic amino acid residues in this region varies from 4 to 0. It would be expected that two consecutive basic amino acids would be present in this region according to the criteria of a bipartite NLS provided by Dingwall and Laskey (1998). The NLS is therefore likely to start with the two basic amino sequence KR and end with the four basic amino acids sequence RR(K/R)K present in all orthologues. These sequences are perfectly aligned and conservation of several other residues within this region is apparent. Furthermore, these predicted NLS sequences meet the criteria of a bipartite NLS, which comprises two clusters of basic amino acids composed of lysine (K) and arginine (R): a smaller cluster of two basic amino acids and a larger cluster of at least three basic amino acids in a group of five. Additionally, for nuclear import to occur, a minimum spacer sequence of 10 residues is required to reach between the two basic clusters of the NLS when these two sites are fitted into the receptor. However, the spacer sequence can be much longer than 10 residues and still direct a protein to the nucleus (Dingwall and Laskey, 1998). The spacer sequence in the proposed region of the bipartite NLS meets this requirement because it contains 15–17 amino acids, rather than 8 or 9 in the alternative region.

Conservation of several short sequences is apparent within the N-termini of StzA proteins (Table 5.9 and Appendix 6). Such sequences include: PRR(R/T), TF(H/S), (S/A)DSG(L/I/V)G(S/T)S(L/I/V), LRD(L/V)E(K/N) and Y(F/I)(L/I/V)DL. These may represent recognition motifs for responses to external stimuli. Regions of proteins that have been conserved during evolution are considered to play important roles in the maintenance of their three-dimensional structure and function (Marsden *et al*., 2002). Thus, the conservation of specific residue types within the primary sequence of StzA may be important in conserving secondary

structures, especially as the conservation of many residues correlates with predicted conserved alpha helices (Appendix 7).

A region of the N-termini of StzA proteins (close to intron 1) appears to be particularly alanine rich in Sordariomycetes (Figure 5.5). Alanine-rich helical regions are characteristic of many DNA-binding proteins and have been implicated in transcriptional repression in a number of proteins, including PacC, which regulates pH in *A. nidulans* (Tilburn *et al*., 1995). The repressor function of alanine rich regions has also been attributed to the *A. nidulans* carbon catabolite repressor CreA, a protein that is known to contain 9 consecutive alanine residues (Dowzer and Kelly, 1991). It is notable that 6 consecutive alanine residues are present in the alanine-rich region of the N-terminus of *G. zeae* StzA (Figure 5.5). *T. reesei* Ace1 is known to act as a repressor of cellulase and xylanase expression (Aro *et al*., 2003), so it could be speculated that the alanine-rich regions within the Ace1/StzA proteins of other Sordariomycetes may contribute to the function of these proteins as transcriptional repressors. *A. nidulans* StzA may also play a role as a repressor, since it has been implicated in the negative regulation of the vacuolar Ca^{2+}/H^{+} exchanger *vcxA*, in addition to its association with cation export from the cell *via enaA* activation (Spielvogel *et al*., 2008).

The lengths and sequences of the C-termini vary considerably among StzA orthologues (Appendix 6). The C-termini vary in size from 69 aa (*V. albo-atrum*) to 363 aa (*A. capsulatus*; Table 5.10). The lack of sequence conservation makes any potential introns within this region hard to predict, and therefore *stzA* orthologues possessing longer C-termini in particular may contain unrecognised introns. It could be hypothesised that the variability of the C-terminal region is responsible for the distinct phenotypes of abiotic stress response and cellulose/xylan utilisation attributed to *A. nidulans* StzA and *T. reesei* Ace1, respectively. Thus, polymorphisms among StzA orthologues, which are from species exhibiting profoundly different lifestyles, may contribute to allowing species to adapt to their particular ecological niches. One clue that the C-terminal region may serve a specific function is the low, but significant homology shown to *rsv1* from *S. pombe*, which encodes a zinc finger TF that is responsive to stress and carbon starvation (Figure 5.6; Hao *et al.,* 1997).

The C-termini of StzA orthologues generally show an over-representation of aspartate and glutamate residues and are particularly rich in proline residues (Table 5.10). These regions may represent ‘acid blob’ and proline rich domains, respectively, which are considered to be involved in initiating transcriptional activation following binding of the zinc fingers to the target DNA (Latchman, 1998). Within this region, and

also the N-termini, each StzA protein contains a number of SPXX and TPXX motifs, which have been found to occur more frequently in regulatory proteins with roles in DNA-binding than in other proteins (Table 5.11; Suzuki, 1989). The *A. nidulans* TFs CreA, AreA and PacC contain high frequencies of these motifs (Dowzer and Kelly, 1991; Kudla *et al*., 1990; Tilburn *et al*., 1995).

5.3.4 Phylogenetic relationships among StzA proteins

Pairs of StzA proteins from species of the same fungal class are generally more similar to each other than proteins from different classes (Table 5.8). However, even among StzA proteins from *Aspergilli*, there are substantial sequence variations. For example, *A. nidulans* StzA is only 32% similar to *A. terreus* StzA, but the StzA proteins of *A. fumigatus* and *N. fischeri* are 95% identical. A phylogenetic analysis of StzA was undertaken to examine these relationships in more detail. However, as exemplified by the contradictions shown by Lutzoni *et al.* (2004) and Liu and Hall (2004), the phylogeny of the Pezizomycotina is still under revision and any such analysis, especially when based upon a single gene, should be approached with caution. One very detailed analysis, which includes most of the organisms described in this analysis, was carried out by Berbee (2001). The phylogeny demonstrated in Figure 5.7a, produced by the analysis of all 29 identified StzA proteins, is broadly in agreement with that of Berbee (2001) and also the phylogenomic investigation of Robbertse *et al.* (2006). Although this does contrast with the work of Fitzpatrick *et al.* (2006) who used the same genome databases to perform a genomic analysis in which *P. nodorum* (a Dothideomycete) clustered more closely with the Sordariomycetes.

Examination of the phylogeny of StzA revealed the clustering of species according to the classes (Eurotiomycetes, Sordariomycetes and Dothideomycetes) and orders (Eurotiales, Onygenales, Sordariales, Hypocreales and Pleosporales) to which they belong. The only exceptions concerned the lack of grouping of two Onygenales (*A. capsulatus* and *P. brasiliensis*) with other members of this class. When the highly variable C-termini of StzA proteins alone were subjected to phylogenetic analysis (Figure 5.7b), a very different picture of relationships was evident. The most striking differences concerned *A. terreus*, *A. oryzae* and *A. flavus* grouping with the Sordariomycetes and the isolation of *A. nidulans* and *A. capsulatus* from the other Eurotiomycetes. This phylogenetic analysis may reflect the evolutionary pressures that these species have encountered that can be related to the function of the C-termini of StzA proteins.

5.3.5 Characterisation of potential TF binding sites in *stzA* promoters

The positions of putative TF binding sites in the promoters (2 kb) of 22 *stzA* orthologues were recorded (Tables 5.12 and 5.13). The extent of sequence divergence among *stzA* promoters became apparent considering that very little sequence conservation was observed when using alignments with ClustalW, Advanced TCoffee and Mlagan, even when restricted to species from a particular class (data not shown). Dramatic divergence of TF binding site motifs in *stzA*/*ace1* promoters (and also StzA/Ace1 target genes) may therefore help to explain why *stzA* from *A. nidulans* and *ace1* from *T. reesei* regulate such distinct phenotypes (abiotic stress tolerance and cellulase/xylanase expression, respectively). Hence, it is important to consider that these phenotypes may not be entirely related to the differences in the amino sequences of the StzA and Ace1 proteins themselves.

Numbers of TF binding site motifs required to be present in promoters to demonstrate significant enrichment for specific TFs were calculated using chi-square analysis (Table 5.14). In the *A. nidulans stzA* promoter, there were potential binding sites for the following TFs: AbaA (7), AlcR (6), AmdA/AmdX (1), AmyR (3), AnCF (3), AreA (4), BrlA (2), CDRE (4), CpcA (1), CreA (10), PacC (1), PecR (3), STRE (2) and XlnR (1). The *A. nidulans stzA* promoter is enriched for the binding site motifs of the TFs AbaA, AmyR and PecR, which control conidiophore development, amylose utilisation and pectin utilisation, respectively (Appendix 9; Andrianopoulos and Timberlake, 1994; Petersen *et al*., 1999; Coutinho *et al*., 2009). Such enrichment of these TF binding motifs may reflect their contribution to *stzA* gene expression.

The discovery of the *REALALE* sequence (5'-CTATCAGGCA) in the *stzA* promoters of 7 *Aspergillus* species indicates an interaction between AreA and StzA in these species (Figure 5.8a). This sequence, 5'-CTATCAGGCA, consists of overlapping putative AreA/StzA binding sites (Wilson and Arst, 1998; O'Neil *et al*., 2002) and is found within the *pol* regions of the *REAL* LTR retrotransposon of *Alternaria alternata* and other filamentous fungi (Kaneko *et al*., 2000; Daboussi and Capy, 2003). Since AreA is the TF that mediates nitrogen metabolite repression in *A. nidulans* (Wilson and Arst, 1998), nitrogen source availability may indicate a further stress to which *stzA* responds. Perturbation of nitrogen source utilisation in *sltA1* mutants and *stzA* deletion strains has already been observed when arginine is present as the sole nitrogen source (Clement *et al*., 1996; O'Neil *et al*., 2002; Chapter 4). Regulation at the *REALALE* sequence may depend on the binding of one TF excluding the binding of the other. It is interesting that although the *REALALE*

sequence is conserved in sequence and position in 11 *stzA* promoters, only one promoter demonstrated enrichment for binding site motifs of AreA and no promoters were enriched for those of StzA (Table 5.15 and Appendix 9). This may indicate that only AreA and StzA binding site motifs that are part of the *REALALE* sequence are functional, which would seem to lend more credence to a successful binding site exclusion regulatory mechanism operating between these two TFs within *stzA* promoters. Binding site exclusion is not an uncommon mechanism for transcriptional regulation. For example, the TF-encoding gene that regulates ethanol utilisation in *A. nidulans*, *alcR*, is autoregulated by AlcR itself and repressed by the carbon catabolite repressor CreA. CreA acts by competing directly with AlcR for binding at the same promoter region (Kulmburg *et al*., 1993). Likewise, the promoter of the *aatA* gene (involved in penicillin biosynthesis) contains overlapping binding sites for AnBH1 and AnCF (the Hap complex). AnBH1 acts as a repressor and counteracts positive regulation by AnCF (Caruso *et al*., 2002).

The presence of *REALALE* in the promoter of the proline transporter-encoding gene *prnB* may indicate that StzA is an additional regulator of this gene, which also contains a putative CpcA binding site in its promoter (Figure 5.9; García *et al.,* 2004). The *REALALE* sequence was found in the promoters of 5 additional genes encoding zinc finger TF proteins (AN5924, AN1893, AN4013, AN8902 and AN8939; Table 5.16), so this sequence may have functional significance for other TFs. *REALALE* was also found within the promoter of a gene encoding a vacuolar ATPase subunit (AN5606) and the promoter of a gene involved in double-strand break repair *via* non-homologous end-joining (AN6705; Table 5.16). These findings may be significant considering that *A. nidulans stzA* is known to be involved in the processes of osmotolerance, cation homeostasis and DNA repair (O'Neil *et al*., 2002; Spielvogel *et al*., 2008; Findon *et al*., 2010).

A CpcA binding site (5'-TGACTCA; Hoffmann *et al*., 2001) is present in 26 out of 29 *stzA* promoters in a conserved position (-55 to -169 upstream from the ATG translation start; Figure 5.8b). This seems especially significant considering that the CpcA binding site motif is expected to occur only once in every 8,192 bp by chance alone (Table 5.14). The *S. cerevisiae* CpcA orthologue, *GCN4*, has been demonstrated to be involved in the regulation of over 500 genes (Natarajan *et al.,* 2001). Given that *cpcA* acts as a central conduit for a wide variety of stress responses, including the signalling of amino acid deprivation in *A. nidulans (*Hoffmann *et al*., 2001), it is perhaps not surprising that CpcA could play a role in the regulation of *stzA*. As the promoter of the *stzA* orthologue in *T. reesei* (*ace1*) is enriched for CpcA binding site motifs (two

sites at -169 and -306 from the ATG translation start), this could indicate a fascinating link between amino acid availability and cellulase gene expression (Kubicek *et al*., 2009).

The AbaA binding site motif (5'-CATTCY; Andrianopoulos and Timberlake, 1994) is significantly enriched in the promoter of *A. nidulans* and also *A. oryzae* and *A. flavus* (Table 5.15 and Appendix 9). AbaA controls conidial development and the *A. nidulans* gene can complement both a conidiation and dimorphic switching defect in a *Penicillium marneffei abaA* mutant (Borneman *et al.*, 2000). This could imply a close association between developmental pathways and *stzA* expression. Another regulator of developmental regulation is *stuA* (AN5836; Dutton *et al*., 1997), and analysis of its promoter revealed enrichment for StzA binding site motifs. Three of the potential 10 StzA binding sites contribute to the triplet StzA binding site motif TGCCTTGCCTTGCCT at -1617 (relative to the ATG translation start site) and may be functional due to their tandem arrangement (Ravagnani *et al*., 1997).

Three *stzA* promoters are significantly enriched for the AlcR binding site motif (5'-CCGCW; Panozzo *et al*., 1997; Appendix 9) compared to none for the control group, suggesting a potential role for this regulator of ethanol utilisation in the expression of *stzA*/*ace1* genes of *P. nodorum, T. reesei* and *A. capsulatus*, and the latter two promoters are also enriched for CreA binding site motifs (5'-SYGGRG; Cubero and Scazzocchio, 1994), indicative of a possible regulatory role for the global carbon catabolite repressor CreA. However, Aro *et al*. (2003) demonstrate that *ace1* in *T. reesei* is unlikely to be subject to CreA-mediated glucose repression.

Eight of the 19 *stzA* promoters analysed (42.1%) were enriched for the AnCF binding site motif 5'-CCAAT (Steidl *et al*., 1999; Table 5.15 and Appendix 9). In *A nidulans*, AnCF serves as a redox sensor in co-ordinating the cellular oxidative stress response (Thön *et al*., 2010), and therefore may regulate *stzA* gene expression in this species in response to oxidative stress to further broaden the stresses to which *stzA* responds. However, *stzA* deletants and *stzA*$^{+}$ strains exhibited similar growth phenotypes in the presence of the oxidative stresses exerted by hydrogen peroxide and menadione (Table 4.37). AnCF is thought to be involved in the regulation of hundreds of genes in *A. nidulans* (Thön *et al*., 2010), so may explain why a relatively high percentage (18.2%) of the control promoters were enriched for AnCF binding site motifs.

5.3.6 Phylogeny and proposed roles of the *REALALE* sequence in *stzA* promoters

The *REALALE* sequence 5'-CTATCAGGCA has been identified in 11 *stzA* promoters, including that of *A. nidulans* (Figure 5.8a). The presence of this conserved sequence within the promoters of *stzA* appears to be linked to class, as species with this sequence are all from the Eurotiomycetes, despite the *REAL* LTR retrotransposon with which it is associated being from the Dothideomycete *A*. *alternata*, for which no *stzA* orthologue sequence data is currently available. It can be speculated that *REALALE* was acquired in *stzA* promoters after the divergence of the Eurotiomycetes from the other classes of the Pezizomycotina containing *stzA* orthologues (Sordariomycetes, Leotiomycetes and Dothideomycetes). It seems that *REALALE* was subsequently lost from some Eurotiomycetes: *A. terreus*, *C. immitis* and the members of the Onygenales Arthrodermataceae family (*A. benhamiae*, *A. gypseum* and *T. rubrum*). It is unknown whether the *T. emersonii stzA* promoter contains the *REALALE* sequence because only 1067 bp of the promoter was available for analysis. *REALALE* sequences were found in promoters between -466 to -975 bp upstream of the ATG translation starts of *stzA* genes, so although possible, it seems unlikely that the *T. emersonii stzA* promoter possesses a *REALALE* sequence.

It seems significant that the majority of species containing *REALALE* within their *stzA* promoters are typically associated with animal pathogenicity (or are non-pathogenic) rather than being linked with plant pathogenicity (Table 5.25). Conversely, organisms lacking the *REALALE* sequence within their *stzA* promoters are generally associated with plant pathogenicity or a saprobic lifestyle, with the exception of *C. globosum*, an organism that is associated with infections in humans, in addition to being a saprobe. Hence, the *REALALE* sequence may have contributed to the ability of *stzA*-possessing fungal species to colonise animal hosts in addition to being linked with phylogeny. Thus, the different roles associated with StzA in *A. nidulans* (a Eurotiomycete) and *T. reesei* (a Sordarimomycete), namely abiotic stress tolerance (O'Neil *et al*., 2002) and cellulose/xylan utilisation (Aro *et al*., 2003), respectively, may reflect not only differences in phylogeny but also the presence and absence of *REALALE* sequences. *REALALE* may represent a transposon-mediated transfer of DNA: a process in which new beneficial TF regulatory sequences were transferred into *stzA* promoters (reviewed in Richards *et al*., 2011). Such sequences may have conferred additional *stzA* phenotypes, such as improved salt tolerance and enhanced DNA repair capabilities to meet the challenges associated with the colonisation of animals or new ecological niches.

Orthologues of *A. nidulans* genes that contained the *REALALE* sequence in their promoters were analysed for the presence of *REALALE* sequences, but no such sequences were found in either *A. fumigatus* or *A. oryzae* promoters (Table 5.16). The *REALALE* sequence may therefore represent an intriguing conserved regulatory mechanism that is specific to *stzA* expression to co-ordinate responses to diverse stresses in particular ecological niches.

Table 5.25. The biological significance of species containing StzA orthologues. Adapted from information available at the Tree of Life Web Project, 2006 (http://tolweb.org/Pezizomycotina/29296) and the Broad Institute (www.broadinstitute.org). Letters beside the name of each species show the class in the Pezizomycotina subphylum to which each species belongs: Eurotiomycetes (E), Sordariomycetes (S), Leotiomycetes (L) and Dothideomycetes (D). Colours represent the fungal lifestyle of each species: animal pathogen (red), occasional animal pathogen (pale red), plant pathogen (green), saprobe (pale green), saprobe and occasional pathogen (brown) and non-pathogen (pale blue). The presence of the *REALALE* sequence (containing potential overlapping binding site motifs for AreA and StzA) within an stzA promoter is indicated by (R) following the name of the species.

	Species	Biological significance
E	*Aspergillus nidulans* (R)	Key fungal model system for genetics and cell biology. Soil saprobe. Rare animal pathogen.
E	*Aspergillus fumigatus* (R)	Major opportunistic pathogen of immunocompromised individuals causing invasive and non-invasive aspergillosis (lung disease) and is also a major allergen. Its main lifestyle is plant decomposition (saprobe).
E	*Aspergilus flavus* (R)	The second most frequent pathogen responsible for invasive and non-invasive aspergillosis (after *A. fumigatus*). The predominant pathogen involved in the production of aflatoxins, potent carcinogens that infect a range of foodstuffs.
E	*Aspergillus oryzae* (R)	Closely related to *A. flavus* but not non-pathogenic. A source of many enzymes including glucoamylase, alpha amylases and proteases used in starch processing, baking and brewing. Used for the preparation of koji in oriental fermented foods.
E	*Aspergillus terreus*	Used to produce organic acids. Opportunistic human pathogen.
E	*Aspergillus niger* (R)	Widely used in the fermentation industry for the production of citric acid. Also can cause aspergillosis as an opportunistic human pathogen and is a common cause of fungal ear infections. Causes disease of onion, peanuts and grapes.
E	*Neosartorya fischeri* (R) (*Aspergillus fischeri*)	Close relative of *A. fumigatus*, but is an extremely rare invasive pathogen.
E	*Aspergillus clavatus* (R)	Close relative of *A. fumigatus*. Occasionally pathogenic but can produce the toxin patulin, associated with disease in animals and humans.
E	*Penicillium chrysogenum* (R)	Non-pathogenic organism used as a source of penicillin.
E	*Uncinocarpus reesii* (R)	Non-pathogenic species but morphologically very similar to *Coccidioides*, which are pathogenic, allowing a comparative approach to the study of pathogenesis in *Coccidioides*.
E	*Ajellomyces capsulatus* (R) (*Histoplasma capsulatum*)	Most common cause of fungal respiratory infections (histoplasmosis), which can be life threatening to the elderly and immunocompromised patients.
E	*Paracoccidioides brasiliensis* (R)	The causal agent of paracoccidioidomycosis, one of the most important systemic mycoses in Latin America.
E	*Coccidioides immitis*	Causes coccidioidomycosis (valley fever), a serious and sometimes fatal disease that can infect healthy people.
E	*Trichophyton rubrum*	The most common dermatophyte species, causing fungal skin infections that include athlete's foot and ringworm.
E	*Arthroderma benhamiae*	Zoophilic dematophyte. Causes intractable superficial infections and triggers severe inflammatory responses. Guinea pigs are natural hosts.
E	*Arthroderma gypseum* (*Microsporum gypseum*)	Soil saprophyte. Occasionally pathogenic to humans causing tinea corporis or tinea capitis.
E	*Talaromyces emersonii*	Thermophilic non-pathogen.
D	*Phaeosphaeria nodorum* (*Stagonospora nodorum*)	Major pathogen of wheat and related cereals.
D	*Pyrenophora tritici-repentis*	Grass pathogen causing significant disease.
S	*Gibberella zeae* (*Fusarium graminearum*)	Member of the *Fusarium genus*, the most important group of fungal plant pathogens, which cause various diseases on almost every economically important plant species.
S	*Magnaporthe grisea*	Major plant pathogen and is the causal agent of rice blast disease. Closely related to *Neurospora crassa*.
S	*Verticillium albo-atrum*	Well-studied phytopathogenic species.
S	*Trichoderma reesei* (*Hypocrea jecorina*)	Saprobic fungus that that efficiently degrades plant cell wall polysaccharides, including cellulose and hemicelluloses.
S	*Podospora anserina*	A saprobic fungus used as a model organism.
S	*Chaetomium globosum*	Important in the decomposition of plant and other cellulose-rich materials. Causative agent of skin and nail infections in humans. Occasionally causes cerebral and systemic infections associated with high mortality rates.
S	*Nectria haematococca*	Colonises a wide variety of habitats. Can be a soil saprobe and a pathogen of plants and animals, including humans.
S	*Neurospora crassa*	Non-pathogen used as a leading model organism for the study of eukaryotic genetics and biology.
S	*Sordaria macrospora*	Model organism for the analysis of fungal development and meiosis.
L	*Botryotinia fuckeliana* (*Botrytis cinerea*)	Widespread phytopathogenic fungus that causes Gray-mould rot or *Botrytis* blight disease, affecting most vegetable and fruit crops and many shrubs, trees, flowers and weeds.

5.3.7 Identification of putative StzA binding sites in potential StzA target genes

StzA (*A. nidulans*) and Ace1 (*T. reesei*) seem to regulate the distinct phenotypes of abiotic stress response and cellulose/xylan utilisation, respectively (O'Neil *et al*., 2002; Aro *et al*., 2003; Chapter 4). The wealth of data available from recently completed genome projects (Haas *et al*., 2011) was used to identify potential TF binding sites within the promoter regions of candidate target genes to assist in the identification of possible gene interactions. This will help to explain the apparent differences in the physiological pathways that these genes control. The conservation of potential regulatory binding sites across different species and genera of organisms is an important aspect of comparative genomics and is used to infer possible TF interactions (Tirosh *et al.*, 2008). The *A. nidulans stzA* gene regulates a wide variety of abiotic stress responses, including osmotic and DNA damage responses, cation transport and pH regulation (O'Neil *et al*., 2002; O'Mahony *et al*., 2002; Spielvogel *et al*., 2008; Findon *et al*., 2010) and is potentially involved in nitrogen metabolism (Sections 5.2.7a and 5.2.7civ). Therefore, promoters of genes involved in these processes in *A. nidulans* (and other *Aspergillus* species where relevant) were analysed for the presence and positional conservation of potential StzA binding sites.

5.3.7a Osmotic stress

The close proximity of two potential StzA binding sites within the *sskB* promoters of the *A. nidulans* (5'-TGCCTGCCT at -1339) and *A. fumigatus* (5'-TGCCTGAGGCA at -1150) may be significant since physiologically relevant TF binding sites are often arranged in tandem (Ravagnani *et al.,* 1997). Hence, StzA may interact with SskB, a MAPKKK of the HOG signalling pathway, which regulates the osmotic stress response (Furukawa *et al*., 2005).

StzA binding sites within the promoters of the *msnA* genes that are part of the sequence 5'-TAGCACAGGCAGATC are conserved in 5 filamentous ascomycetes (*A. nidulans, A. fumigatus, N. fischeri, A. clavatus* and *A. terreus*) and thus may also have functional significance. It has been shown that *msnA* is expressed in response to a variety of stress-related stimuli, including hydrogen peroxide and heat stress, in addition to high salt concentrations (Han and Prade, 2002). Hence, *msnA* is not exclusive to the HOG

pathway but is a general stress response TF activated by other MAPKs or stress pathways. This would appear to raise the possibility that StzA interacts with *msnA*.

StzA may also play a role in the regulation of the histidine kinases *tcsA* and *phkB* since the corresponding promoters are both enriched for StzA binding site motifs (9 occurrences in each promoter; Table 5.17). Histidine kinases mediate signal transduction in response to stimuli. Only 4 of the 15 genes encoding histidine kinases in the *A. nidulans* genome have been characterised (*tcsA*, *tcsB*, *fphA* and *nikA*; Suzuki *et al*., 2008; Miskei *et al*., 2009). The *tcsA* gene in *A. nidulans* is required for sporulation under normal osmotic conditions (Virginia *et al*., 2000).

5.3.7b DNA repair

It is notable that the promoter of the AN6705 gene, which contains the *REALALE* sequence (at -846; Table 5.16), also contains the potential double StzA binding site sequence 5'-TGCCTGAGGCA (at -316) which may bind StzA due to its tandem arrangement (Ravagnani *et al.,* 1997). This gene is a component of the RSC chromatin remodelling complex and is the orthologue of *S. cerevisiae RSC8* involved in double-strand break repair *via* non-homologous end-joining. This process is used to repair DNA when homologous DNA sequences are not available as a template (reviewed in Goldman and Kafer, 2004). StzA is known to function in DNA repair, with *sltA1* mutants and *stzA* deletion strains exhibiting sensitivity to DNA-damaging agents (O'Neil *et al*., 2002; Chapter 4). Hence, StzA may interact with the AN6705 gene to repair double-stranded DNA *via* non-homologous end-joining. Since a CpcA binding site (at -1179) is present in a similar region to the *REALALE* sequence, this promoter may bind CpcA in addition to enabling AreA/StzA interactions.

Multiple potential StzA binding sites are present in conserved positions in the promoters of the *ku80* gene in a region approximately -60 to -150 bp from the ATG translation start sites (Figure 5.10). This gene in *A. nidulans* (AN4552), like AN6705, is involved in the repair of double-stranded DNA *via* non-homologous end-joining (Goldman and Kafer, 2004). StzA binding site motifs appear to be over-represented in *ku80* promoters, especially in *Aspergillus* species, which may be indicative of a regulatory role attributable to StzA (Table 5.19). However, it should be noted that *ku80*$^{+}$ and *ku80*$^{-}$ *stzA* deletion strains responded similarly to

DNA-damaging agents analysed (Figure 4.28). Hagiwara *et al*. (2009) identified a consensus sequence of TGAC(G/A)TCA in the promoters of SskA- HogA- and AtfA-dependant fludioxonil up-regulated genes. The reverse complement of this sequence (highlighted in grey in Figure 5.10) is conserved in all 7 *Aspergillus ku80* promoters for which sequences were available. Furthermore this sequence is remarkably similar to the cAMP response element (CRE) motif, 5'-T(G/T)ACGT(C/A)A, to which AtfA and Sko1p are able to bind. Hence, it is possible that either or both of these proteins may regulate *ku80* gene expression along with StzA. AtfA has been shown to regulate general stress responses in *A. nidulans* by interacting with HogA (Lara-Rojas *et al*., 2011).

5.3.7c pH regulation and cation homeostasis

The promoter of *A. nidulans pacC* is enriched for StzA binding site motifs (Tables 5.14 and 5.20), with 2 of these 8 motifs overlapping with potential PacC binding sites, indicating that StzA may influence autoregulation of the *pacC* gene by competing for PacC binding sites. This could be highly significant considering that PacC is the principal pH-responsive TF in *A. nidulans* (Tilburn *et al*., 1995). Moreover, it has been shown that StzA is necessary for growth of *A. nidulans* at alkaline pH under buffered conditions (Spielvogel *et al*., 2008). The presence of double conserved StzA binding sites within the *palA* promoters of 5 *Aspergillus* species (-98 to -160 from the 5' ATG translation start sites) may indicate a role of StzA in the regulation of *palA*, which participates in pH signal transduction in *A. nidulans* (Negrete-Urtasun *et al*., 1997; Figure 5.11).

Analysis of *stzA* deletion mutants has revealed their sensitivity to alkaline pH and also to Mg^{2+}, Li^{+} and Cs^{+} (Spielvogel *et al*., 2008) in addition to their sensitivity to Na^{+} and K^{+} (Spathas, 1978), implicating a wider role for StzA in cation tolerance. The binding of StzA to the promoters of both *enaA* (positive regulation) and *vcxA* (negative regulation) has been determined experimentally (Spielvogel *et al*., 2008), but it is unknown whether StzA regulates the transcription of other ion transporters. The study demonstrated that StzA regulates expression of *ENA1*, which encodes a homologue of the *S. cerevisiae* P-type Na^{+} ATPase and pumps Na^{+} out of the cell during conditions of high extracellular Na^{+}. Binding of StzA to *enaA* has been confirmed experimentally, and yet the *enaA* promoter has only 3 putative StzA binding sites (Table 5.21). This is fewer than would be expected by chance alone (3.91 occurrences per 2 kb promoter). StzA was also shown to act

negatively on transcription of *vcxA*, which encodes a homologue of the *S. cerevisiae* vacuolar Ca^{2+}/H^{+} exchanger, in conditions of high extracellular Ca^{2+}. The *vcxA* gene has only 2 potential *stzA* binding sites in its promoter.

It is thought that *pmrA* (AN7464), a P-type golgi apparatus Ca^{2+}/Mn^{2+} ATPase, is likely to act as the major Ca^{2+} ATPase during salt stress because increased expression of both *stzA* and *enaA* has been observed in a *pmrA* deletion strain of *A. fumigatus* (Pinchai *et al*., 2010). These findings indicate an interaction of StzA and EnaA with *pmrA*. Although *A. nidulans pmrA* has only one StzA binding site (Table 5.21), this may be functional.

Spielvogel *et al*. (2008) have identified two further orthologues of the *S. cerevisiae* ENA sodium ATPase system in *A. nidulans*: *enaB* and *enaC*, known to be involved in the salt stress response and are required for growth at alkaline pH. Although this research group was unable to induce expression either of these two genes in 300 mM LiCl or 100 mM $CaCl_2$, they remain potential candidates for the regulation of the salt stress response in *A. nidulans*, and possess 4 and 7 potential StzA binding sites in their promoters, respectively (Table 5.21). Hence, it is possible that StzA, in addition to AtfA, SsrA and RcoA (Miskei *et al*., 2009) targets the *enaC* gene.

The *pmcA* and *pmcB* genes in *A. nidulans* encode vacuolar calcium pumps that show increased transcript levels in an *sltA* deletion mutant (Findon *et al*., 2010), indicating negative regulation of these genes by StzA. The *pmcA* gene has only 3 potential StzA binding sites in its promoter, whereas the promoter of *pmcB* has 10 potential StzA binding sites and is therefore significantly enriched for this motif (Tables 5.14 and 5.21).

5.3.7d Nitrogen metabolism

The presence of conserved putative doublet StzA binding sites within the promoters of *cpcA* genes indicates that StzA may have a role in the regulation of *cpcA*, which mediates the cross pathway control of amino acid biosynthesis (Hoffmann *et al*., 2001; Table 5.23 and Figure 5.13). It is noteworthy that physiologically relevant TF binding sites are often arranged in tandem or multiple organisation (Ravagnani *et al.,* 1997). Evidence for interaction between StzA and *cpcA* was evident in both Eurotiomycetes and Sordariomycetes,

yet *A. nidulans* StzA and *T. reesei* Ace1 TFs control quite different phenotypic responses. Hence, these two proteins may interact with *cpcA* in a manner that has evolved to report different stress responses dependent on their variable C-termini. Transcriptional autoregulation is known to occur for *cpcA (*Hoffmann *et al*., 2001), and is a process that appears to be restricted to filamentous fungi (Braus *et al*., 2006). Due to the close proximity of the conserved putative doublet StzA binding site to the CpcA binding region (~140 bp upstream) within *cpcA* orthologues and its filamentous specific nature, it is possible that StzA is a candidate for interfering with *cpcA* autoregulation. Additionally, StzA binding site motifs appear over-represented in many *cpcA* promoters (Figure 5.14). The presence of conserved StzA binding sites within the promoters of *A. nidulans* and *A. fumigatus cpcB*, which acts antagonistically to *cpcA*, may indicate StzA–*cpcB* interactions further regulate the stress response (Figure 5.12; McCahill *et al*., 2002). Furthermore, it was evident that a CpcA binding site (5'-TGACTCA; Hoffmann *et al*., 2001) was present in 26 out of 29 *stzA* promoters in a conserved position (-55 to -169 of the ATG translation starts; Figure 5.8b). This indicates that CpcA may bind to *stzA* promoters in addition to the possibility of StzA binding to *cpcA* promoters. Due to the Pezizomycotina specific nature of *stzA* and its orthologues, and that *cpcA* plays a key role in fungal pathogenesis *in A. fumigatus* (Krappmann *et al*., 2004), StzA may present itself as a target for the development of antifungal agents.

5.3.7e Statistical analysis of StzA binding site motifs within the promoters of *A. nidulans* genes assigned to functionally related groups

Chi-square analysis of the frequency distribution of potential StzA binding site motifs within the promoters of genes involved in the osmotic stress response and genes containing the *REALALE* sequence in their promoters indicates that both groups may be enriched with functionally significant StzA binding sites. Calculated chi-square values are well above the critical values for both groups of genes when compared to a control group of genes (Table 5.24; Appendix 10). Moreover, the frequency distribution of potential StzA binding sites is clearly skewed to the right for promoters containing the *REALALE* sequence when compared to other groups of genes, including the control group (Figure 5.15). This indicates that promoters containing the *REALALE* sequence are endowed with functional StzA binding sites. Promoters of genes involved in the DNA damage response may also be enriched with functionally significant StzA binding sites as the calculated chi-square value is around the level of significance. Promoters of genes involved in cation

homeostasis are unlikely to be enriched with functionally significant StzA binding sites as calculated chi-square values are well below the critical values.

It is important to interpret the results with consideration to the following error sources:

- Some genes in the test groups may not be under direct StzA control but regulated indirectly by a second TF.

- Some genes in the control group may be under StzA control. For example, the gene encoded by AN7250 is an Na^+/H^+ exchanger, which may be regulated by StzA since StzA is known to be involved in cation homeostasis (Appendix 8).

- Sizes of promoter regions are an estimate. Promoter regions of 2 kb upstream of the ATG translation start sites were analysed. However, Coutinho *et al*. (2009) analysed promoter regions of just 1 kb. The analysis of larger promoter regions (2 kb) might reduce the significance of the results if functional binding sites are closer to the ATG translation start sites, but the analysis of 1 kb promoter regions might have resulted in the exclusion of functional binding sites.

- There are instances where even functional StzA binding sites within promoter regions might not contribute to the expression of the gene analysed, but instead contribute to the expression of an adjacent gene on the complementary DNA strand.

- Greater numbers of StzA binding site motifs in promoters are indicative of an increased likelihood that functional StzA binding sites are present. However, some CreA and PacC target genes are known to possess promoters containing fewer binding site motifs for these TFs than would be expected by chance alone (Espeso and Peñalva, 1996; Hynes *et al*., 2002).

5.3.8 Conclusion

The manual annotations required for *stzA*/StzA orthologues seem to reflect the fact that annotation programs are lagging behind in their ability to accurately pinpoint coding regions compared to the vast increases in sequence data that have recently become available (Wortman *et al*., 2009; Haas *et al*., 2011). Following manual annotation, 15 (out of 29) and 19 (out of 22) of the available *stzA* genes and their corresponding proteins differed in sequence at NCBI and the Broad Institute, respectively (Tables 5.4–5.6). Orthologues of *stzA* genes appear to be limited to filamentous ascomycetes and are further restricted to the Pezizomycotina, a subphylum containing species associated with both plant and animal pathogenicity (Tables 5.1–5.3 and 5.25). Because StzA proteins are restricted to a single subphylum, they may present themselves as attractive targets for the development of antifungal compounds while avoiding adverse effects on their mammalian and plant hosts. This strategy seems feasible considering that the emergence of microbial drug resistance has resulted in a shift from attempts to kill micro-organisms to targeting virulence factors that hinder their ability to cause harm to the host (reviewed in Gauwerky *et al*., 2009).

Conservation of the putative CpcA binding site 5'-TGACTCA (Hoffmann *et al*., 2001) within promoters of *stzA* orthologues (26 out of 29) and the conserved putative doublet StzA binding site within 11 *cpcA* promoters are indicative of *stzA-cpcA* interactions that co-ordinate stress responses with amino acid biosynthesis (Table 5.23 and Figures 5.8b, 5.13). Intriguingly, the promoter of the *stzA* orthologue in *T. reesei* (*ace1*) was found to be significantly enriched for CpcA binding sites (two sites at -169 and -306 from the ATG starts; Table 5.13 and Appendix 9), a finding that could indicate a link between amino acid availability and cellulase gene expression (Kubicek *et al*., 2009). Furthermore, the presence of the *REALALE* sequence (5'-CTATCAGGCA), which contains putative overlapping binding sites for AreA and StzA, within the *stzA* promoters of 11 species from the Eurotiomycetes class, including 7 *Aspergillus* species, indicates an interaction between StzA and AreA, the global regulator of nitrogen metabolite repression (Figure 5.8a; Wilson and Arst, 1998). These findings would seem especially significant considering that StzA plays a key role in the osmotic stress response in *A. nidulans* (O'Neil *et al*., 2002; Spielvogel *et al*., 2008) and a physiological association of the HogA pathway with nitrogen metabolism has been demonstrated for this species *(*Hagiwara *et al*., 2009). Several nitrate/nitrite transporters and a nitrate reductase were downregulated in HogA-dependent manner when exposed to fludioxonil or osmotic stress. Furthermore, promoters of genes involved in the osmotic stress response in *A. nidulans* were significantly enriched for

StzA binding site motifs compared to the control group, indicating that StzA may regulate many of these genes in the HogA pathway (Table 5.24). *REALALE*-containing promoters of other *A. nidulans* genes were also found to be significantly enriched for StzA binding site motifs (Table 5.24), but the implications of this finding remain to be elucidated.

It is clear that promoter sequences have diverged significantly among all *stzA* orthologues (Table 5.13), so their associated phenotypes are likely to exceed any explanations that can be accounted for solely by the presence or absence of *REALALE*. Since StzA proteins demonstrate high variability in the lengths (69–363 aa) and sequences of their C-termini, it is possible to speculate that polymorphisms within these regions are also likely to account substantially for phenotypic differences among orthologues (Table 5.10; Appendix 6).

Chapter 6

General Discussion

6.1 The pleiotropic nature of *A. nidulans* StzA

O'Neil *et al.* (2002) described the cloning of a gene encoding a C_2H_2 zinc finger protein (StzA) that alleviated sensitivity to a variety of abiotic stresses in an *A. nidulans sltA1* mutant. The *sltA1* mutation was originally recognised as conferring sensitivity to Na^+ and K^+ (Spathas, 1978). Since then, it has been shown that *sltA1* strains are sensitive to high arginine concentrations (>10 mM) when present as a sole nitrogen source (Clement *et al*., 1996). Complementation of the *sltA1* mutation in GO281 with the *stzA* gene to yield transformants STC1–4 indicated that StzA also alleviates sensitivity to DNA-damaging agents (O'Neil *et al.*, 2002). Although STC1 and 2 fully complemented sensitivities to salt, arginine, the alkylating agent MNNG, UV irradiation and the UV mimic 4NQO, STC3 did not complement MNNG sensitivity and STC4 did not complement salt, MNNG or 4NQO sensitivities. The different levels of *sltA1* complementation seen in STC3 and 4 may have been due to ectopic and multiple integration events of *stzA* that might have reduced the apparent level of *sltA1* complementation. Moreover, as StzA is a TF, gene transcription in these transformants might have been altered at a variety of loci. The discrepancies in strains STC3 and 4 highlighted the need to demonstrate allelism of *sltA* with *stzA* and to verify the phenotypes controlled by StzA.

This aim was achieved in the present study by deletion of the *stzA* gene and subsequent phenotypic analyses. The phenotypes of *stzA* deletion strains with regards to cation sensitivity (Na^+, K^+, Mg^{2+}, Li^+, Al^{3+} and Rb^+), arginine sensitivity (50 mM as an N-source) and DNA repair sensitivity (UV, 4NQO and MNNG) were similar in intensity to those of GO281 (Figures 4.25, 4.28 and 4.35). These results implied that *stzA* is allelic with *sltA* and the identity of its sequence could be confirmed by observing sequence data from the *A. nidulans* genome project (Galagan *et al*., 2005). To avoid confusion, the gene is referred to as "*stzA*" throughout the discussion.

During the course of the present study, an *stzA* deletion strain has been constructed by Spielvogel *et al*. (2008) and showed sensitivity to a number of cations, the antibiotic neomycin and alkaline pH. However, the study did not test whether *stzA* deletion confers arginine sensitivity and DNA damage sensitivity upon this strain, and therefore did not confirm that these two important pleiotropic effects can be attributed to StzA. Furthermore, *stzA* deletion mutants in the present study exhibited growth phenotypes similar to wild-type strains at pH 8.0, indicating that StzA does not have a regulatory role in adaptation to alkaline pH in

unbuffered MM (Figure 4.36), unlike the results obtained by Spielvogel *et al*. (2008) at this pH for buffered MM. The finding that *stzA* deletion mutants are sensitive to the cations Al^{3+} and Rb^{+} in the present study is novel (Figure 4.25).

It had been hypothesised that *sltA1* encodes an abnormally salt-sensitive arginase enzyme that is allelic to the arginase gene *agaA* (Clement *et al*., 1996). This hypothesis was supported by the close proximity of the *sltA1* mutation and the *agaA* locus, which are shown to be adjacent on linkage group VI (Clutterbuck, 1997). However, examination of the *A. nidulans* genome sequence within this region revealed that it is highly unlikely that *agaA* and *stzA* are allelic as these genes are separated by over 50 kb, encompassing 12 known or predicted genes (Barham-Morris, 2006). Recent BLAST searches at NCBI show no domain matches of StzA with arginase. The present study, however, does reveal that the salt sensitivity and arginine sensitivity phenotypes are likely to be both controlled by defects in the *stzA* gene as *stzA* gene deletion strains are sensitive to both cations and high concentrations of arginine (50 mM) when provided as a sole nitrogen source (Figures 4.25 and 4.35). Co-segregation of arginine and salt sensitivity phenotypes and co-restoration of wild-type phenotypes was always observed in recombinant progeny resulting from sexual crosses between *sltA1* and *sltA*$^{+}$ strains, consistent with a mutation in a single gene (Figure 3.3 and Appendix 3). Hence, in the presence of high arginine concentrations, it seems likely that StzA interacts directly with *agaA,* or indirectly *via* its specific activator ArcA (Borsuk *et al*., 1999; Empel *et al*., 2001). It has been hypothesised that the guanidinium moiety of arginine competes with Na^{+} for the Na^{+} binding site within the Na^{+}/H^{+} antiporter (Attwell *et al*., 1995). Another explanation for the apparent role of *stzA* in *agaA* expression is co-expression of these genes. Genes in close proximity on the chromosome tend to be positively co-expressed more than would be expected to occur by chance because open chromatin makes adjacent genes amenable to transcription, even if they are not immediate neighbours (Batada *et al*., 2007). Thus, common TF binding sites are not necessary for co-expression. This phenomenon might explain why *agaA* was induced in the presence of salt stress in the absence of arginine in L20, but was only partially induced in GO281 (Barham-Morris, 2006).

The reason for the involvement of a TF, such as StzA, in regulating tolerance to both DNA-damaging agents and salt is unknown. However, it is known that salt/osmotic stress induces both sporulation and secondary metabolite production (Harris *et al*., 2009). Since many secondary metabolites mutate the structure of nucleic acids, it has been speculated that the simultaneous induction of DNA repair/stabilising enzymes is necessary

(Paterson and Lima, 2008). This is particularly relevant to *Aspergillus* species that produce aflatoxins, which are the most mutagenic natural compounds produced (Yu *et al*., 2004a). Hence, it would not be surprising for conditions of osmotic stress in *A. nidulans* to co-induce secondary metabolite production and DNA repair mechanisms. Furthermore, fungal phytopathogens/saprophytes trigger a hypersensitive response upon plant infection, producing a desiccating environment with DNA-damaging agents present (reviewed in Gururani *et al*., 2012; Mannuss *et al*., 2012). In this context, the ability of a single TF, such as StzA, to co-ordinate both osmotic stress and DNA repair responses confers a selective advantage. Proteins with roles in DNA repair might well evolve different functions. A dual-function protein with roles in DNA repair and sporulation has already been identified in *A. nidulans*. CryA acts as both a photolyase, which repairs UV-induced DNA damage, and a cryptochrome (a blue-light receptor that regulates sexual development; Bayram *et al*., 2008).

Adding to its role as a stress response regulator, StzA has recently been associated with calcium homeostasis (Findon *et al*., 2010). *sltA1* deletion strains demonstrated highly elevated transcript levels for *pmcA* and *pmcB*, which encode vacuolar ATPases, and *vcxA,* which encodes a vacuolar membrane Ca^{2+}/H^{+}exchanger. These changes were associated with hypertrophy of the vacuolar system.

6.2 Characterisation of the *sltA1* mutation in GO281

The *sltA1* mutation in *A. nidulans* GO281 is known to lack a (completely) functional *stzA* gene due to the presence of a point mutation at nucleotide 2962 (G>A), which introduces a premature termination codon (PTC) at amino acid 502 (W>X; O'Neil *et al*., 2002). The resultant mutant protein will therefore have lost its entire C-terminal domain, immediately following the zinc finger region, which includes a domain rich in proline residues that is thought to have a role in transcriptional activation. The phenotypic similarities of *stzA* deletion strains and GO281 showing similar severe sensitivity to cations, DNA-damaging agents and arginine (Figures 4.25, 4.28 and 4.35) indicate that the *sltA1* mutation in GO281 *stzA* is a total loss-of-function mutation, i.e. no StzA protein is translated in GO281. If a partly functional StzA protein were translated in GO281, it seems reasonable to expect that this strain would be less sensitive to these stresses compared to the *stzA* deletion strains. Furthermore, Northern analysis has indicated that *stzA* mRNA accumulation in GO281 is consistently dramatically reduced compared to L20 (Barham-Morris, 2006).

Lack of a functional StzA protein in GO281 could be due to mRNA instability arising from nonsense mediated decay (NMD) of the truncated mRNA transcript. NMD is the regulated degradation of mRNA containing a PTC. This process is a major checkpoint for transcript fidelity that prevents cellular damage from potentially harmful proteins resulting from abnormalities in gene expression. PTC-induced transcript degradation is associated with mRNA 3′ tagging and dissociation of mRNA from ribosomes. The *nmdA1*, *cutA* and *cutB* genes regulate this process in *A. nidulans*. NMD is also known to be involved in the degradation of many wild-type mRNA transcripts in this organism (Morozov *et al*., 2006; Morozov *et al*., 2012).

Even if a truncated StzA protein in GO281 were translated and the zinc fingers still able to bind DNA, it is likely that the lack of a transcriptional activating domain would result in ineffective autoregulation and protein instability or an altered folding pattern that would distort the protein to reduce its functionality (Shen, 2010). An *in silico* comparison of the characteristics of full-length StzA and truncated GO281 StzA, by Vector NTI software, predicted StzA protein instability of truncated StzA due to major differences between both the altered charge at pH 7 and the isoelectric point compared with wild-type StzA (Barham-Morris, 2006).

Defective carbon source utilisation phenotypes have previously been attributed to the *sltA1* mutation. Such speculation arose due to the poor growth of GO281 observed on glycerol and ethanol as sole carbon sources when compared to the strong growth of the wild-type strain L20 (O'Neil *et al*., 2002). Furthermore, the present study revealed that GO281 grows poorly on a wide range of carbon sources, many of which are substrates of the gluconeogenic pathway (Table 3.1 and Figure 3.1). However, *stzA* deletion strains grew equally well as L20 and control strains on representative carbon sources of the gluconeogenic and glycolytic pathways and also on the plant substrates starch, xylan and xylose (Figures 4.29, 4.30 and 4.32). These results indicate that the defect in carbon source utilisation in GO281 is not due to the *sltA1* mutation but due to a second uncharacterised mutation(s), and therefore StzA has no regulatory role in carbon metabolism. This assertion is supported by phenotypic analyses of recombinant progeny resulting from sexual crosses between *sltA1* and *sltA*$^{+}$ strains, where defective carbon source utilisation phenotypes segregated independently of salt, arginine and DNA damage sensitivity phenotypes (Figure 3.3 and Appendix 3).

6.3 Distribution and functions of StzA proteins

Twenty-nine StzA proteins have been identified by BLAST searches at NCBI and the Broad Institute (Table 5.5). Numerous StzA annotation errors were evident, resulting mainly from missed introns, which were predicted manually using protein and DNA alignments, intron positional conservation and known fungal intron structure. Following manual annotation, 16 (out of 29) and 19 (out of 22) of the available *stzA* genes and their corresponding proteins differed in sequence at NCBI and the Broad Institute, respectively (Tables 5.4–5.6). The extensive manual annotations required for StzA orthologues appear to reflect the fact that gene annotation programs are lagging behind in their ability to accurately pinpoint coding regions from the wealth of sequence data that has recently become available (Wortman *et al*., 2009; Haas *et al*., 2011).

All StzA orthologues are restricted to filamentous ascomycetes of the Pezizomycotina subphylum (Chilton *et al*., 2008). This subphylum contains highly diverse fungal species that serve as model organisms, saprophytes, plant pathogens and animal pathogens that are important for basic research, health, agriculture and industry, as well as for comparative genomics (reviewed in Machida and Gomi, 2010; Table 5.25). Because StzA proteins are restricted to a single subphylum, they may present themselves as attractive targets for the development of antifungal compounds that avoid adverse effects on their mammalian and plant hosts. This strategy seems feasible considering that the emergence of microbial drug resistance has resulted in a shift from attempts to kill micro-organisms to targeting virulence factors that hinder their ability to cause harm to the host (reviewed in Gauwerky *et al*., 2009; Hartmann *et al*., 2011).

StzA proteins belong to the C_2H_2 class of TFs and are highly conserved across the regions containing the zinc fingers and putative nuclear localisation signal. By contrast, few residues are conserved within the N-termini of StzA proteins but may have important roles in the maintenance of secondary structure, considering their correlation with conserved alpha helices predicted for this region (Appendix 7). Regions of proteins that have been conserved during evolution are considered to play important roles in the maintenance of both their structure and function (Marsden *et al*., 2002). Hence, the conservation of specific residue types within the primary sequence of StzA may be important for not only conserving secondary and tertiary protein structures but also functioning as recognition motifs for responses to external stimuli.

Surprisingly, the C-termini of StzA proteins are highly variable in both length (69 to 363 aa from the first residue after the second H residue of the third zinc finger to the end of the proteins) and sequence, with no apparent conservation of individual residues or secondary structure (Appendices 6 and 7). This may explain why the phylogenetic analysis results of the C-termini are very different from those obtained using full-length StzA proteins and the N-termini alone (Figure 5.7). Full-length StzA proteins and the N-termini appear to demonstrate phylogenetic relationships that truly reflect the evolutionary histories of their respective genomes (Berbee, 2001; Robbertse *et al*., 2006; Wang *et al*., 2009), whereas the numerous polymorphisms in the C-termini might reflect adaptation of a species to its particular ecological niche. Hence, it is the sequence variations of the C-termini that are thought most likely to influence the divergent functions of StzA proteins (Chilton *et al*., 2008). Accordingly, it could be hypothesised that the variability of the C-terminal region is responsible for the distinct phenotypes of abiotic stress response and cellulose and xylan utilisation attributed to *A. nidulans* StzA and *T. reesei* Ace1, respectively (O'Neil *et al*., 2002; Aro *et al*., 2003). One clue that the C-terminal region may serve a specific function is the significant homology shared between *A. nidulans stzA* and *S. pombe rsv1*, which encodes a zinc finger TF that is responsive to stress and carbon starvation (Figure 5.6; Hao *et al.,* 1997). The inability of functional domains to be assigned to the N- and C-termini of selected StzA proteins by Conserved Domain Database (CDD) searches at NCBI and the Broad Institute is evidence that StzA and its orthologues are a novel group of proteins that have yet to be characterised, with the exception of StzA from *A. nidulans* and Ace1 from *T. reesei*.

The *T. reesei ace1* mutant (VTT-D-061244) and wild-type strain (QM9414) exhibited similar phenotypic responses to the osmotic stresses NaCl and KCl; the DNA-damaging agents UV, 4NQO and MNNG; high arginine concentrations (50 mM) provided as a sole nitrogen source; neomycin; and acidic and alkaline pH (Figure 4.39). The results indicate that Ace1 from *T. reesei* is not a regulator of stress responses but has a role restricted to cellulase and xylanase expression. By contrast, *A. nidulans* StzA appears to have no role in cellulase, xylanase or pectinase/polygalacturonase expression (Table 4.8). These results are further evidence that StzA and Ace1 have evolved to regulate very different phenotypic responses despite sharing 58% overall amino acid similarity (Aro *et al*., 2003).

6.4 Predicted interactions of *stzA* with nitrogen metabolic regulators

The discovery of the *REALALE* sequence in eleven *stzA* promoters belonging to the Eurotiomycetes class, including the *Aspergillus* species examined (except *A. terreus*), indicates an interaction between AreA and StzA in these species (Figure 5.8a). This hypothesis is based on the *REALALE* sequence (5'-CTATCAGGCA) comprising overlapping putative AreA/StzA binding sites (Wilson and Arst, 1998; Spielvogel *et al*., 2008). Since AreA is the TF that mediates nitrogen metabolite repression in *A. nidulans* (Wilson and Arst, 1998), nitrogen source availability may indicate a further stress to which *stzA* responds. Perturbation of nitrogen source utilisation in *sltA1* mutant and *stzA* deletion strains has already been observed when arginine is present as the sole nitrogen source (Clement *et al*., 1996; O'Neil *et al*., 2002; and the present study). Transcriptional regulation at the *REALALE* sequence may depend on binding of one TF excluding binding of the other. AreA might interact with *stzA* during conditions of stress when nitrogen sources are limiting. Conversely, in the absence of these stresses, StzA might bind to its own promoter in an autoregulatory mechanism that suppresses transcription of StzA and prevents binding of AreA.

As the *REALALE* sequence is found within the *REAL* LTR retrotransposon of *Alternaria alternata* and other filamentous fungi (Kaneko *et al*., 2000; Daboussi and Capy, 2003), this sequence within *stzA* promoters may have been derived from transposon-mediated transfer (Chilton *et al*., 2008). Intriguingly, Tf1, a long terminal repeat retrotransposon in *S. pombe*, is known to have a strong preference for promoters that are induced by conditions of stress (Guo and Levin, 2010). Hence, it was hypothesised that Tf1 is either directed to or better retained in the promoters of stress response genes to enhance survival of *S. pombe* cells exposed to environmental stresses, including conditions of DNA damage (MMS) and osmotic stress (sorbitol). Furthermore, this mechanism of retrotransposon integration into the promoters of stress response genes could extend to other organisms to counter threats to their survival. Examples that detail how transposons have evolved to specifically react to stress are numerous (reviewed in Beauregard *et al*., 2008; Richards *et al*., 2011). It is quite probable that transposon-mediated transfer has introduced new beneficial TF binding sites into *stzA* promoters to enable rapid change by selection of new target genes, ultimately enabling species to colonise additional ecological niches. This mechanism may account for the differing functions associated with *stzA* in *A. nidulans* (contains *REALALE*) and *ace1* in *T. reesei* (lacks *REALALE*), which appear to regulate the distinct phenotypes of abiotic stress tolerance and cellulose and xylan utilisation, respectively (O'Neil *et al*., 2002; Aro *et al*., 2003). The proposed *REALALE* sequence may have regulatory

significance that extends to other *A. nidulans* genes since *REALALE*-containing promoters in this species were found to be significantly enriched for StzA binding site motifs, indicating regulation of these genes by StzA (Figure 5.15 and Table 5.24).

Apart from the *REALALE* sequence, promoter alignments revealed a distinct lack of nucleotide sequence conservation in *stzA* promoters. A CpcA binding site (5'-TGACTCA; Hoffmann *et al*., 2001), however, was present in 26 out of 29 *stzA* promoters in a conserved position (-55 to -169 bp upstream from the ATG translation start site; Figure 5.8b). This seems especially significant considering that the CpcA binding site motif is expected to occur only once in every 8,192 bp by chance alone (Table 5.14). The *S. cerevisiae* CpcA orthologue, *GCN4*, has been demonstrated to be involved in the regulation of over 500 genes (Natarajan *et al.,* 2001). Given that *cpcA* acts as a central conduit for a wide variety of stress responses, including the signalling of amino acid deprivation in *A. nidulans (*Hoffmann *et al*., 2001), it is feasible that CpcA could play a role in the regulation of *stzA* (Chilton *et al*., 2008). Furthermore, a conserved putative doublet StzA binding site (5'-TGCCTN$_{(12\text{-}18)}$TGCCT) has been identified in 11 *cpcA* promoters (Table 5.23 and Figure 5.13). Since physiologically relevant binding sites are often conserved in position between orthologues and/or arranged in tandem or multiple organisation (Ravagnani *et al.,* 1997), this further indicates that *stzA*-*cpcA* interactions might co-ordinate stress responses with amino acid biosynthesis.

The interaction of CpcA with *stzA* may not be surprising considering that StzA plays a key role in the osmotic stress response in *A. nidulans* (O'Neil *et al*., 2002; Spielvogel *et al*., 2008) and a physiological association of the HOG pathway with nitrogen metabolism has been demonstrated for this species *(*Hagiwara *et al*., 2009). Several nitrate/nitrite transporters and a nitrate reductase were downregulated in a HogA-dependent manner when exposed to fludioxonil or osmotic stress.

The presence of two statistically significant CpcA binding sites in the promoter of *T. reesei ace1* (Table 5.13; Appendix 9) could indicate an intriguing link between amino acid availability and cellulase gene expression (Kubicek *et al*., 2009). Ace1 is known to regulate cellulase expression (Aro *et al*., 2003) and methionine has been shown to increase cellulase expression in *T. reesei* (Gremel *et al*., 2008). Hence, proposed CpcA interactions with *ace1* could potentially be exploited in the development of strains with high expression levels of secreted cellulases to improve upon the productivity of similar strains currently used in the pulp and paper, food and textile industries and for the production of sugars and bio-ethanol (Kubicek *et al*., 2009).

6.5 A proposed model for StzA as a novel regulator of general stress responses

The *A. nidulans* genome contains genes orthologous to all those of the high osmolarity glycerol (HOG) mitogen-activated protein kinase (MAPK) pathway of *S. cerevisiae*, which is involved in stress signalling, particularly osmoadaptation (Furukawa *et al*., 2005; Miskei *et al*., 2009). In *A. nidulans*, both the osmotic and oxidative stress responses are regulated by this general stress signalling pathway, in which AtfA interacts with the MAPK SakA (HogA). HogA accumulates in the nucleus in response to stress signals where it interacts with AtfA (Balázs *et al*., 2010; Lara-Rojas *et al*., 2011). The HOG pathway is also involved in transcriptional responses to the fungicide fludioxonil (Hagiwara *et al*., 2009). Deletion of components of the *A. nidulans* HOG pathway does not cause extreme sensitivity to high osmolarity, unlike interception of this pathway in *S. cerevisae* (Furukawa *et al*., 2005). This implies that stress signalling is more complex in *A. nidulans* compared to yeast and that an unknown osmoresponsive pathway(s) still remains to be identified (Furukawa *et al*., 2005; Etxebeste *et al*., 2010).

It is notable that although *A. nidulans* StzA is involved in tolerance to ionic osmotic stresses and the fungicide neomycin (O'Neil *et al*., 2002; Spielvogel *et al*., 2009), StzA was not identified as a component of the HOG pathway in the study by Miskei *et al*. (2009), which involved a comprehensive annotation of stress-response proteins in *Aspergilli*. Nor have more recent studies shown StzA to be a component or target gene of the HOG pathway. Hence, StzA would appear to be a key candidate as a component of the proposed alternative osmoresponsive pathway. Furthermore, promoters of genes involved in the osmotic stress response in *A. nidulans* were significantly enriched for StzA binding site motifs when compared with the control group, indicating that StzA may regulate many of these genes in the HOG pathway (Table 5.24). Moreover, with a central role in general stress responses in *A. nidulans*, StzA regulates tolerance to a wide array of stresses, including not only osmotic stress, DNA-damaging agents, high arginine concentrations and neomycin, but also cation, calcium and pH homeostasis, and inferably nitrogen source availability (Spathas, 1978; Clement *et al*., 1996; O'Neil *et al*., 2002; Chilton *et al*., 2008; Spielvogel *et al*., 2008; Findon *et al*., 2010; Section 6.4). A proposed model for StzA in the regulation of general stress responses that is independent of the HOG pathway is shown in Figure 6.1. StzA does not appear to regulate tolerance to sources of non-ionic osmotic stress (Figure 4.27), oxidative stress (Figure 4.37) or the fungicide fludioxonil (Figure 4.38) – all of which are known to activate the *A. nidulans* HOG pathway (Hagiwara *et al*., 2009). Nor

does StzA seem to regulate the Cell Wall Integrity (CWI) pathway, a major regulator of cell wall biogenesis resulting from environmental stresses that include hypo-osmolarity and heat shock (Table 4.38; Levin, 2011).

The major components of the *A. nidulans* HOG pathway are known to be conserved among all other *Aspergillus* species that have been sequenced, indicating the presence of very similar regulatory stress signalling processes among *Aspergilli* (Miskei *et al*., 2009). However, there is a notable absence in *Aspergillus* species of important HOG pathway transcriptional regulators (including Sko1p, Hot1p and Msn1p) and low homologies of the *Aspergillus* proteins YpdA, MsnA, AtfA, NapA, HsfA and HacA compared to their counterparts in yeast. Furthermore, no StzA orthologues have been identified in yeast. Hence, stress response mechanisms are likely to be more similar among *Aspergillus* species than between any one *Aspergillus* species and yeast. This could indicate that StzA is part of a common stress regulatory mechanism conserved among *Aspergillus* species, but absent in yeast, considering that StzA orthologues have been identified in all *Aspergillus* species sequenced.

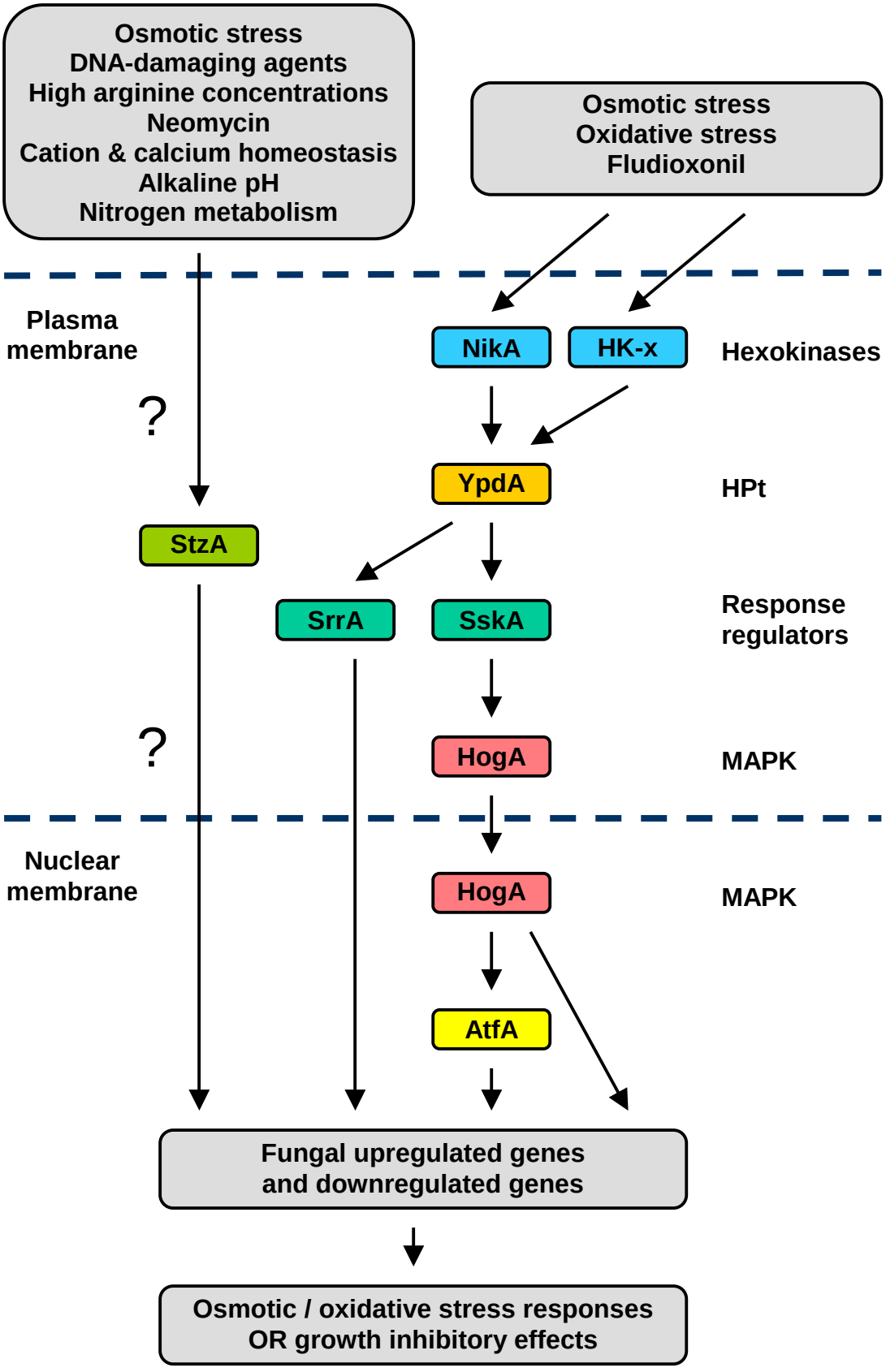


Figure 6.1. A proposed model for the regulation of general stress responses in *A. nidulans* by the HOG pathway and an unknown pathway involving StzA. The model for His–Asp phosphorelay signalling and the HogA(SakA)–AtfA MAPK cascade in *A. nidulans* was proposed by Hagiwara *et al*. (2009) and Lara-Rojas *et al*. (2011). The hexokinase (HK) NikA serves as a sensor that responds to osmotic stress, oxidative stress and the fungicide fludioxonil. NikA transmits the signal *via* YpdA (histidine-containing phosphotransfer intermediate; HPt) to SskA and SrrA (response regulators; RR). The signal is then transduced to the MAPK HogA, which localises to the nucleus where it interacts with AtfA to initiate a transcriptional response.

6.6 StzA as a potential antifungal drug target

Infections caused by opportunistic fungal pathogens have become a major clinical problem during the past few decades due to increasing numbers of immunocompromised patients. This has largely been due to increases in patients who have been diagnosed with HIV and AIDS, increases in organ and bone marrow transplant patients undergoing immunosuppressive therapy and increases in cancer patients treated with cytotoxic drugs (Morschhäuser, 2010). *A*. *fumigatus* is one of the key pathogens responsible for these fungal infections and is the main causal agent of invasive aspergillosis. The mortality rate in most aspergillosis patients is approximately 50% despite recent advances in chemotherapy (Espinel-Ingroff, 2009). Existing antifungals are often toxic or ineffective because of resistance, so additional drug targets are needed to develop new antifungal drugs (Gauwerky *et al*., 2009).

Stress response pathways in *A. fumigatus* are being considered as promising candidates for novel and specific antifungal drug targets (Hartmann *et al*., 2011). The resistance of *A. fumigatus* to stressful environments is considered to have contributed substantially to its development as a human pathogen, considering the stresses it encounters when invading a susceptible human host. These stresses include elevated temperature, alkaline pH, oxidative stress, hypoxia, nutritional deprivation, extensive secretion and cell membrane perturbations during treatment with antifungals. StzA in *A. fumigatus* may represent a promising antifungal drug target if this TF is shown to regulate tolerance to a range of stresses in a similar manner to its orthologue in *A. nidulans*. StzA also has the advantage of being Pezizomycotina-specific, so represents a drug target that could minimise any adverse effects on humans. Furthermore, *stzA* potentially interacts with CpcA, which has been shown to be necessary for virulence in mice infected with pulmonary aspergillosis (Krappmann *et al*., 2004). This has led to the recent suggestion that the scope of the "CPCome" should be determined (Hartmann *et al*., 2011). The findings of the present study indicates that it is highly likely that *stzA* is part of the "CPCome" in *A. fumigatus* and other filamentous ascomycetes (Section 6.4).

A comparison of the roles of *A. fumigatus* and *A. nidulans* StzA proteins in stress adaptation may establish the potential of *A. fumigatus* StzA as a valid antifungal drug target. This aim could be achieved by amplifying the *stzA* gene from *A. fumigatus*, along with its native promoter, and transforming into an *A. nidulans stzA* deletion strain followed by phenotypic and expression analyses. Such a strategy avoids the need for elaborate safety measures and animal models by using a safe laboratory expression system for an *A.*

fumigatus TF. Hence, the relevance of *A. fumigatus* StzA to stress responses could be screened in a non-pathogenic background, while exploiting the extensive legacy of physiological, genetical and molecular techniques of *A. nidulans*.

6.7 Suggested future investigations

It is unknown whether promoters containing TF binding site sequences evolve to accommodate regulation by TFs or whether TF proteins themselves evolve to regulate new target genes. Both types of changes are likely to contribute to evolution but their relative contributions are unknown (reviewed in Wagner and Lynch, 2008). The actions of *stzA* orthologues are likely to be influenced by variations in: (a) StzA protein sequences, (b) *stzA* expression due to binding sites for differing TFs in *stzA* promoters, (c) StzA target genes (determined by functional StzA binding sites in target promoters). Some of these factors could be investigated to answer questions arising from the current work.

With regards to changes in StzA protein sequences, the highly variable C-terminal regions are thought most likely to account for the differing phenotypes attributed to StzA and Ace1 proteins in *A. nidulans* and *T. reesei*, respectively. Hence, exploring the unknown functions of the C-termini of StzA and Ace1 seems a worthwhile priority for future investigations, and could be achieved using techniques already established in our laboratory; namely, conventional PCR, fusion PCR, fungal transformation, phenotypic plate tests and DNS enzyme assays. For example, the C-terminus of *T. reesei ace1* could be fused to a DNA fragment containing the promoter, N-terminus and zinc finger region of *A. nidulans stzA*. Likewise, the C-terminus of *A. nidulans stzA* could be fused to a DNA fragment containing the promoter, N-terminus and zinc finger region of *T. reesei ace1.* The chimeric genes could be used to transform the parental strains, and if they are transcribed and translated, expression analysis could be used to detect any differences in cation, DNA damage and arginine sensitivities compared with cellulase and xylanase expression for *stzA*/*ace1* deletion and wild-type strains.

Variation in transcriptional regulation is also considered to be a major contributor of phenotypic variation. Gene regulation resulting from altered TF binding has been shown to be a major cause of phenotypic divergence across related yeast species (Borneman *et al*., 2007). However, it is not always clear what

constitutes the DNA binding site in protein–DNA interactions. A putative DNA binding site may not be functional and many TFs are able to tolerate sequence variation in DNA target sites, to which they may bind with varying affinities under differing environmental conditions (Zheng *et al*., 2010). For example, the mammalian TF AP1 (which is homologous to *GCN4* from *S. cerevisiae* and CpcA) is able to tolerate single nucleotide substitutions and binds to TGACTCA variants with affinities that are physiologically relevant (Seldeen *et al*., 2009). Chromatin structure also affects the accessibility of TF binding sites (Tirosh *et al*., 2008; Wittkopp, 2010).

Although the present study has provided important clues regarding *stzA* gene/protein interactions, it is unknown which TFs bind to the *stzA* promoter and to which promoters StzA binds. Are the CpcA binding sites in *stzA* promoters functional? Does the *REALALE* sequence contain functional AreA and StzA binding sites? Do any of the promoters involved in the osmotic stress response or containing the *REALALE* sequence, and significantly enriched for StzA binding sites, contain StzA binding sites that are functional?

A multi-discipline approach involving techniques such as transcriptome studies using microarrays and the development of high throughput proteomics studies, which are still in their infancy in *Aspergilli*, will aid greatly to our understanding of TF interactions (Andersen and Nielsen, 2009). Experimental validation of functional binding sites for CpcA, AreA and StzA in *stzA* promoters and other candidate promoters could be obtained using techniques involving chromatin immunoprecipitation (ChIP). ChIP-seq (ChIP followed by sequencing) and ChIP-chip (ChIP followed by "chip"; microarray) analyses not only allow all binding sites for a particular TF to be identified and compared across related genomes but also provide estimates of the relative binding affinities for each sequence under different conditions (reviewed in Wittkopp, 2010). Hence, the promoters to which StzA binds could be identified and compared for each *Aspergillus* species under a range of stresses, with relative binding affinities providing clues to the strength and relevance of interactions. Such profiles, allowing comparisons of gene expression with StzA binding, would provide valuable insights into the roles of StzA proteins among species of this genus of considerable medical and biotechnological importance.

In practice, microarray technology is restricted to a few research groups. A conventional quantitative PCR (qPCR) approach could be used by our research group to study the induction of genes in an *A. nidulans stzA* deletion mutant compared to a wild-type strain under selected stresses to begin to pinpoint the genes under

StzA control (Sambrook and Russell, 2001). Phenotypic and/or bioinformatic analyses presented in the present study have indicated that suitable genes to investigate are: those involved in the HOG pathway (*hogA*, *atfA*, *msnA, sskB*, *tcsA*, *phkB* and *enaC*), NER repair genes (including AN8713 and AN0604), alkylation repair genes (including AN2276 and AN4587) and the major regulators of the cross pathway control of amino acid biosynthesis (*cpcA* and *cpcB*).

It should be realised that post-transcriptional and/or post-translational mechanisms may contribute substantially to the regulation of StzA, e.g., by the regulated decay or enhanced stability of *stzA* transcripts in addition to their rate of synthesis. Regulated mRNA decay is known to be involved in the degradation of several wild-type mRNAs in higher eukaryotes but this process is poorly undertood in fungi (Maquat, 2005; Morozov *et al*., 2006). However, in *A. nidulans* it is known that regulated transcript degradation mediated by RrmA is involved in the regulation of arginine metabolism by influencing the levels of *agaA* and *otaA* transcripts, and it has been speculated that such regulation may generally contribute to nitrogen metabolite repression in *A. nidulans* (Olszewska *et al*., 2007). The extent to which the regulated decay or enhanced stability of wild-type mRNA transcripts contributes to post-transcriptional regulatory mechanisms in *A. nidulans* remains to be elucidated (Morozov *et al*., 2012).

The two candidate nuclear localisation signals KRSLSAASTDEGVQRSMARRKK and RRKKNAPPMNINKKCK in the N-terminus of *A. nidulans* StzA could be tested by deletion of these motifs followed by fusion of StzA with green fluorescent protein (GFP). This would allow the subcellular location of StzA to be tracked under different StzA-inducing conditions to determine if either sequence is essential for nuclear entry (reviewed in Lorang *et al*., 2001).

Over 100 fungi have been sequenced or are being sequenced as part of the Fungal Genome Initiative (www.broadinstitute.org). Complete genome sequences of eight *Aspergillus* species have now been determined (Andersen and Nielsen, 2009; Wortman *et al*., 2009). As the number of sequenced genomes for *Aspergillus* species increases, more studies based on comparative genomics should provide further insights regarding the mechanisms of stress regulation in *Aspergilli* and on the role of StzA in this process. Comparative in-depth gene expression studies of filamentous fungi will be facilitated by online resources such as the Filamentous Fungal Gene Expression Database (FFGED; Zhang and Townsend, 2010). This database was recently established to provide a platform for depositing and retrieving data from gene

expression studies specific to filamentous fungi and combines experimental results from diverse types of wet-lab experiments and *in silico* procedures.

6.8 General conclusions

- Generation and phenotypic analyses of *stzA* gene deletion strains of *Aspergillus nidulans* implies that *stzA* is allelic to *sltA*, with the encoded transcription factor regulating tolerance to cations, DNA-damaging agents and high arginine concentrations. The similar severe sensitivity of a *sltA1* mutant (GO281) and *stzA* deletion mutants to these stresses indicated that the premature termination codon in *sltA1* represents a total loss-of-function mutation. It was also verified that StzA has no regulatory role in the utilisation of carbon sources. Findings were supported by phenotypic analyses of recombinant progeny resulting from sexual crosses between *sltA1* and *sltA*$^{+}$ strains.

- Bioinformatic analysis of genes involved in the osmotic stress response revealed that their promoters were significantly enriched for StzA binding site motifs compared to those of the control group, indicating that StzA may regulate many of these genes that comprise the HOG pathway. Although this pathway is activated by fludioxonil, *stzA* deletants and *stzA*$^{+}$ strains showed similar sensitivities to this fungicide. Phenotypic analyses indicate that StzA does not regulate tolerance to sources of oxidative stress, non-ionic osmotic stress or components of the CWI pathway.

- *A. nidulans* StzA appears to have no role in cellulase or xylanase expression as revealed by the results of a DNS assay. *T. reesei ace1* deletant and wild-type strains showed similar sensitivities to cations, DNA-damaging agents, arginine, neomycin, acidic and alkaline pH. These results confirm that *A. nidulans* StzA and *T. reesei* Ace1 regulate the distinct phenotypes of abiotic stress tolerance and cellulase and xylanase expression, respectively, despite these two proteins sharing 58% overall amino acid similarity.

- All twenty-nine StzA orthologues identified are restricted to filamentous ascomycetes of the Pezizomycotina subphylum and may therefore represent specific and novel antifungal drug targets. The C-termini of StzA proteins are highly variable in both length and sequence, with no apparent

conservations in amino acids or predicted secondary structure. This region is considered most likely to influence the divergent functions of StzA proteins. Conservations of individual residues in the N-termini correspond to conserved secondary structure (alpha helices) among StzA proteins, implying shared functions for StzA proteins in this region.

- Regulators of two major nitrogen metabolic pathways (CpcA and AreA) may regulate *stzA* expression. Statistically significant putative CpcA binding sites were positionally conserved in 26 out of 29 *stzA* orthologue promoters, indicating an interaction between *stzA* and CpcA, a transcription factor that mediates the cross pathway control of amino acid biosynthesis. *REALALE* sequences, likely to be of retrotransposon origin, containing putative overlapping binding sites for StzA and AreA, were found in eleven *stzA* promoters of the Eurotiomycetes class, indicating an interaction between *stzA* and the global nitrogen metabolite repressor AreA. Intriguingly, *REALALE*-containing promoters identified across the genome of *A. nidulans* were significantly enriched for StzA binding site motifs when compared to a control group of genes. Hence, *REALALE* may have regulatory significance that extends to other *A. nidulans* genes.

Appendices

Appendix 1 – Genotypes of strains and plasmids

Organism	Strain	Genotype
Aspergillus nidulans	L20	*wA3 pabaA1*
	GO281	*yA2 pabaA1 sltA1*
	L19	*yA2 pyroA4*
	G191	*pyrG89 pabaA1 fwA1 uaY9*
	A1149	*pyrG89 pyroA4 nkuA::argB*
	WV101	*yA2 pyrG*$^{+}$ *sltA1*
	WV202	*yA2 pyrG*$^{+}$ *sltA*$^{+}$
Trichoderma reesei	QM9414	wild-type
	VTT-D-061244	*ace1*
Escherichia coli	JM109	*end*A1, *rec*A1, *gyr*A96, *thi, hsd*R17 (rk–, mk+), *rel*A1, *sup*E44, Δ(*lac-pro*AB), [F´ *tra*D36, *pro*AB, *laq*IqZΔM15].

Plasmid	Genotype	Note
pDEL1	*pyr4*	Described by Nielsen *et al*. (2006)
pDJB3	*pyr4*, *ans1*	Described by Ballance and Turner (1985)
pUC19	*bla, lacZ, rep*	Genbank accession L09137

Appendix 2 – Media and reagents

All media and reagents were autoclaved at 10 psi for 20 minutes, except for solutions containing MOPS, ethanol, isopropanol and other organic solvents, which were filter sterilised.

Malt Extract (ME) broth (per L)

20 g malt extract
pH adjusted to 6.5 with NaOH

Malt Extract Agar (MEA)

ME broth with the addition of 1.5% agar

***Aspergillus* minimal medium (per L)**

10 g glucose
1.84 g ammonium tartrate (10 mM)
20 ml *Aspergillus* salt solution
10 ml *Aspergillus* vitamin solution
20 ml pyrimidine stock solution added for pyrimidine auxotrophic strains
1 ml paba stock solution added for paba auxotrophic strains
50 µl pyro stock solution added for pyro auxotrophic strains
pH adjusted to 6.5 with 1 M NaOH

***Aspergillus* salt solution (per L)**

26 g KCl
26 g $MgSO_4.7H_2O$
76 g KH_2PO_4
50 ml trace element solution

***Aspergillus* trace element solution (per L)**

40 mg sodium borate
400 mg copper sulphate
800 mg ferrous sulphate
800 mg magnesium sulphate
8 g zinc sulphate
Stored at 4 ºC

***Aspergillus* vitamin solution (per L)**

80 mg para-amino benzoic acid (paba)
80 mg inositol
20 mg nicotinic acid
120 mg calcium-D-pantothenate
50 mg pyridoxine
20 mg riboflavin
280 mg choline chloride
400 mg putrescine
2 ml biotin stock solution (10 mg biotin in 100 ml SDW)
Stored at 4 ºC

Pyrimidine stock solution (per 100 ml)

5.6 g uracil
6.1 g uridine

Para-amino benzoic acid (paba) stock solution (per 100 ml)

1 g paba

Pyridoxine (pyro) stock solution (per 100 ml)

0.1 g pyro

MNNG stock solution (per 10 ml)

125 mg MNNG (*N*-Methyl-*N*'-Nitro-*N*-Nitrosoguanidine)

4NQO stock solution (per 10 ml)

30 mg 4NQO (4-nitroquiniline oxide)

TLC tank buffer (per 100.5 ml)

40 ml toluene
60 ml ethyl acetate
0.5 ml formic acid

Sterigmatocystin standards (per ml)

5 µg or 100 µg sterigmatocystin in 1 ml benzene

Luria Bertani (LB) broth (per L)

16 g tryptone
10 g yeast extract
5 g sodium chloride
20 mM magnesium sulphate (9.78g $MgSO_4.7H_2O$) added for the growth of competent cells
pH adjusted to 7.0

Luria Bertani Agar (LBA)

LB broth with the addition of 1.5% agar
LBA plates with ampicillin contained 100 µg ml^{-1} ampicillin

TFB1 (per L)

2.93 g potassium acetate (30 mM)
2.19 g calcium chloride (10 mM)
9.88 g manganous chloride (50 mM)
12.10 g rubidium chloride (100 mM)
150 g glycerol (15%)
pH adjusted to 5.8 using acetic acid

TFB2 (per 100 ml)

0.21 g PIPES (10 mM)
1.64 g calcium chloride (75 mM)
0.12 g rubidium chloride (10 mM)
15 g glycerol (15%)
pH adjusted to 6.5 using acetic acid/sodium hydroxide

Buffer P1 (resuspension buffer; per L)

50 mM Tris-HCl, pH 8.0 (6.06 g Tris)
10 mM EDTA (3.72 g $Na_2EDTA.2H_2O$)
100 µg ml^{-1} RNaseA (100 mg)
pH adjusted to 8.0 with HCl
Stored at 4 ºC

Buffer P2 (lysis buffer; per L)

200 mM NaOH (8.0 g)
1% SDS (50 ml 20% SDS)

Buffer P3 (neutralisation buffer; per L)

3.0 M potassium acetate (294.5 g)
pH adjusted to 5.5 with glacial acetic acid

Buffer QBT (equilibration buffer; per L)

750 mM NaCl (43.83 g)
50 mM MOPS (10.46 g), pH 7.0
15% isopropanol (150 ml)
0.15% Triton® X-100 (15 ml)
pH adjusted to 7.0 with NaOH

Buffer QC (wash buffer; per L)

1.0 M NaCl (58.44 g)
50 mM MOPS (10.46 g), pH 7.0
15% isopropanol (150 ml)
pH adjusted to 7.0 with NaOH

Buffer QF (elution buffer; **per L)**

1.25 M NaCl (73.05 g)
50 mM Tris-HCl, pH 8.5 (6.06 g Tris)
15% isopropanol (150 ml)
pH adjusted to 8.0 with HCl

5% SDS stock solution (per 100 ml)

5 g SDS (sodium dodecyl sulphate)

1 M Tris-HCl stock solution (per 100 ml)

12.1 g Tris
pH adjusted to 8.0 with 1 M HCl

10 mM EDTA stock solution (per 100 ml)

0.372 g disodium EDTA (ethylene diamine tetra-acetic acid)

Extraction buffer (per 100 ml)

10 ml SDS stock solution (0.5%)
20 ml Tris-HCl stock solution (200 mM)
1.46 g NaCl (250 mM)
0.93 g EDTA (25 mM)

TE buffer – 10 mM Tris-HCl, pH 8.0, 1 mM EDTA (per 100 ml)

1 ml Tris-HCl stock solution
10 ml 10 mM EDTA stock solution
pH adjusted to 8.0 with HCl

Tris-Borate EDTA (TBE) buffer (per L)

10.78 g Tris (0.089 M)
5.50 g Boric acid (0.089 M)
0.74 g EDTA (2 mM)

TBE agarose gels (per 100 ml)

1x TBE buffer (100 ml)
0.8 g agarose
50 µl ethidium bromide stock solution (final concentration 0.5 µg ml^{-1})

Ethidium bromide stock solution (per 100 ml)

100 mg ethidium bromide

6x Gel-loading buffer (per 50 ml)

0.125 g bromophenol blue (0.25%)
0.125 g xylene cyanol (0.25%)
20 g sucrose (40%)
Stored at 4 ºC

Buffer QG (binding and solubilisation buffer)

5.5 M guanidine thiocyanate
20 mM Tris-HCl, pH 6.6

Buffer PE (elution buffer)

10 mM Tris-HCl, pH 7.5
80% ethanol

0.6 M KCl protoplasting buffer (per 200 ml)

8.95 g KCl

0.6 M KCl protoplasting buffer with 50 mM $CaCl_2$ (per 200 ml)

8.95 g KCl
1.11 g $CaCl_2$

0.6 M KCl protoplasting buffer with Lysing enzyme (per 15 ml)

0.67 g KCl
300 mg Lysing enzyme (Sigma: L1412)

PEG solution – 25% PEG 6000, 0.6 M KCl, 50 mM $CaCl_2$, 10 mM Tris-HCl, pH 7.5 (per 200 ml)

50 g PEG (polyethylene glycol)
8.95 g KCl
1.11 g $CaCl_2$
0.242 g Tris
pH adjusted to 7.5 with 1 M HCl

Regeneration media base (per L)

Aspergillus minimal medium
1 ml paba stock solution
50 µl pyro stock solution
44.75 g KCl
1.5% agar

Regeneration media overlay

Same as regeneration media base but with only 0.5% agar

Xylan substrate solution (per 50 ml)

0.5 g xylan added to 30 ml 0.2 M acetate buffer, pH 4.5
Made up to 50 ml with SDW

Cellulose substrate solution (per 50 ml)

0.5 g cellulose added to 30 ml 0.2 M acetate buffer, pH 4.5
Made up to 50 ml with SDW

Polygalacturonic acid substrate solution (per 50 ml)

0.5 g polygalacturonic acid added to 30 ml 0.2 M acetate buffer, pH 4.5
Made up to 50 ml with SDW

Sodium acetate buffer 0.1 M, pH 4.5 (per 100 ml)

0.35 ml acetic acid
0.32 g sodium acetate
2 g NaCl
Dissolved in 50 ml SDW
Made up to 100 ml with SDW

Dintrosalicylic acid (DNS) solution (per 100 ml)

1 g DNS
1.6 g sodium hydroxide pellets
30 g potassium sodium tartrate
Dissolved in 50 ml SDW
Made up to 100 ml with SDW
Stored at ambient temperature for up to 10 weeks

Lactose solution (per 100 ml)

Must be freshly prepared
0.0120 g of lactose monohydrate dissolved in 100 ml SDW

DNS-lactose solution (per 100 ml)

Must be freshly prepared
3 parts DNS solution (75 ml) added to 1 part lactose solution (25 ml)

Xylose standard solution (per 100 ml)

0.5 g xylose in 100 ml SDW

Glucose standard solution (per 100 ml)

0.5 g glucose in 100 ml SDW

Galacturonic acid standard solution (per 50 ml)

0.5 g galacturonic acid in 50 ml SDW

50 mM potassium phosphate buffer, pH 7.8 (per 100 ml)

0.72 g K_2HPO_4
0.115 g KH_2PO_4

Bovine Serum Albumin (BSA) stock solution (per 400 ml; 25 µg ml^{-1})

0.010 g BSA
Made up to 400 ml with 50 mM potassium phosphate buffer, pH 7.8

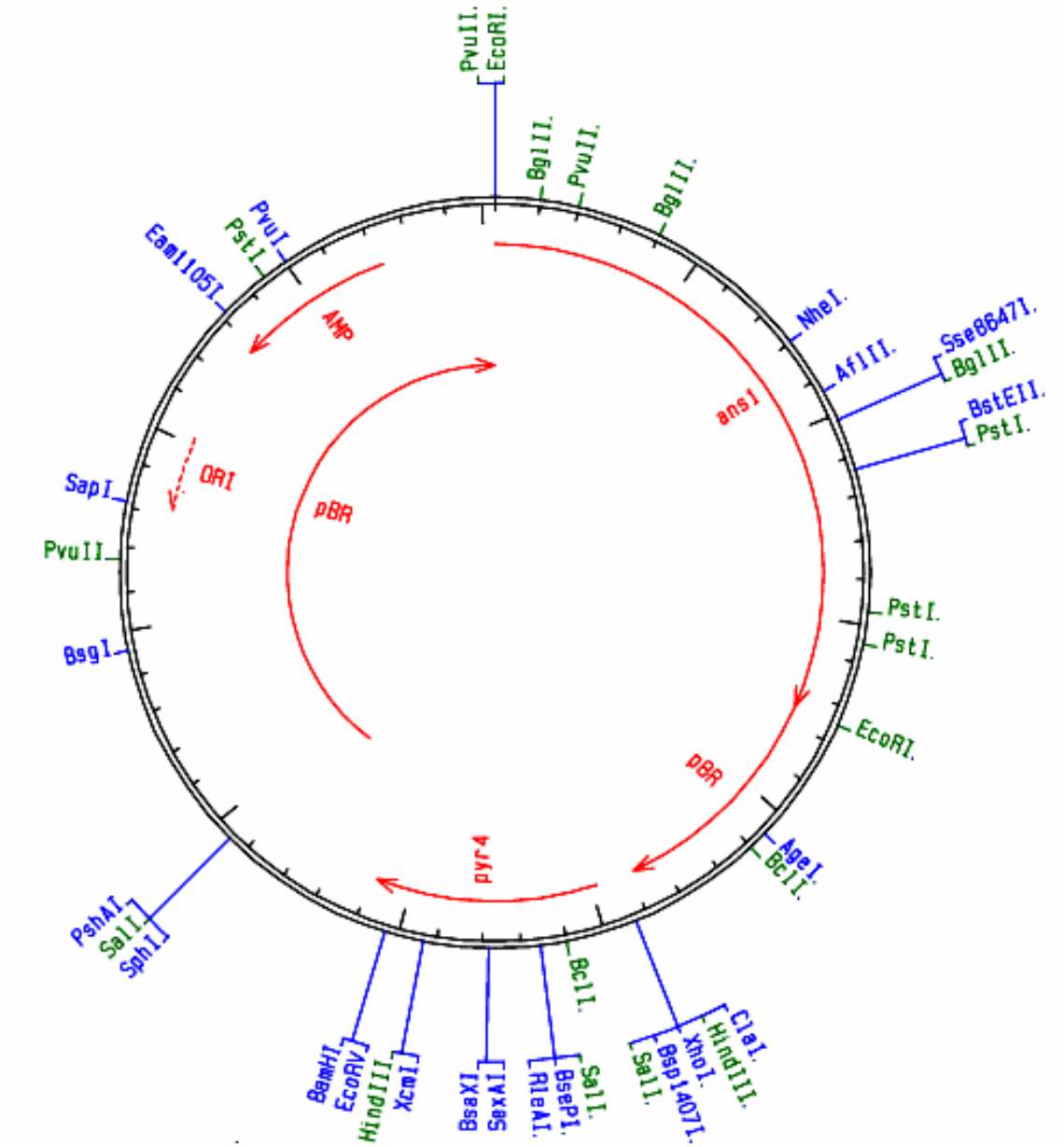
Appendix 3 – Segregation of vitamin, carbon source utilisation and stress markers in 100 recombinant progeny from a sexual cross between GO281 and WP1

Key	G = Good / normal growth
	X = No Growth
	P = Poor / reduced growth

	Vitamin				Carbon source / stress treatment					
Progeny	paba pyro	paba	pyro	Neither vitamin	Glycerol	Ethanol	Galacturonic acid	0.5 M NaCl	UV Irradiation (6 mins)	50 mM Arginine (N-source)
GO281	G	G	X	X	P	P	P	P	P	P
WP1	G	X	G	G	G	G	G	G	G	G
1	G	X	X	X	P	P	P	G	G	G
2	G	G	G	G	G	G	G	G	G	G
3	G	X	G	X	P	P	P	P	P	P
4	G	G	G	G	G	G	G	G	G	G
5	G	X	G	X	G	G	G	G	G	G
6	G	G	G	G	P	P	P	G	G	G
7	G	X	G	X	P	P	P	G	G	G
8	G	X	G	X	P	P	P	G	G	G
9	G	G	X	X	G	G	G	P	P	P
10	G	G	G	G	P	P	P	P	P	P
11	G	X	G	X	G	G	G	G	G	G
12	G	G	G	G	P	P	P	P	P	P
13	G	G	X	X	P	P	P	G	G	G
14	G	G	G	G	G	G	G	G	G	G
15	G	G	G	G	P	P	P	P	P	P
16	G	G	G	G	G	G	G	G	G	G
17	G	X	X	X	P	P	P	P	P	P
18	G	G	X	X	G	G	G	P	P	P
19	G	X	X	X	P	P	P	P	P	P
20	G	X	G	X	P	P	P	P	P	P
21	G	G	G	G	G	G	G	P	P	P
22	G	G	X	X	G	G	G	G	G	G
23	G	X	X	X	G	G	G	G	G	G
24	G	G	G	G	P	P	P	G	G	G
25	G	X	G	X	G	G	G	G	G	G
26	G	X	X	X	P	P	P	G	G	G
27	G	G	X	X	P	P	P	G	G	G
28	G	G	G	G	P	P	P	P	P	P
29	G	G	G	G	G	G	G	P	P	P
30	G	G	G	G	P	P	P	G	G	G
31	G	X	G	X	G	G	G	G	G	G
32	G	X	G	X	P	P	P	P	P	P
33	G	G	G	G	G	G	G	G	G	G
34	G	G	G	G	P	P	P	P	P	P
35	G	G	G	G	G	G	G	G	G	G
36	G	X	G	X	G	G	G	P	P	P
37	G	X	X	X	P	P	P	G	G	G
38	G	G	G	G	P	P	P	G	G	G
39	G	X	G	X	G	G	G	G	G	G
40	G	X	G	X	G	G	G	G	G	G
41	G	X	X	X	G	G	G	G	G	G
42	G	X	G	X	G	G	G	P	P	P
43	G	X	G	X	G	G	G	G	G	G
44	G	X	X	X	G	G	G	P	P	P
45	G	X	G	X	G	G	G	G	G	G
46	G	X	X	X	G	G	G	G	G	G
47	G	X	G	X	P	P	P	P	P	P
48	G	G	G	G	G	G	G	G	G	G
49	G	G	X	X	G	G	G	P	P	P
50	G	G	X	X	P	P	P	P	P	P

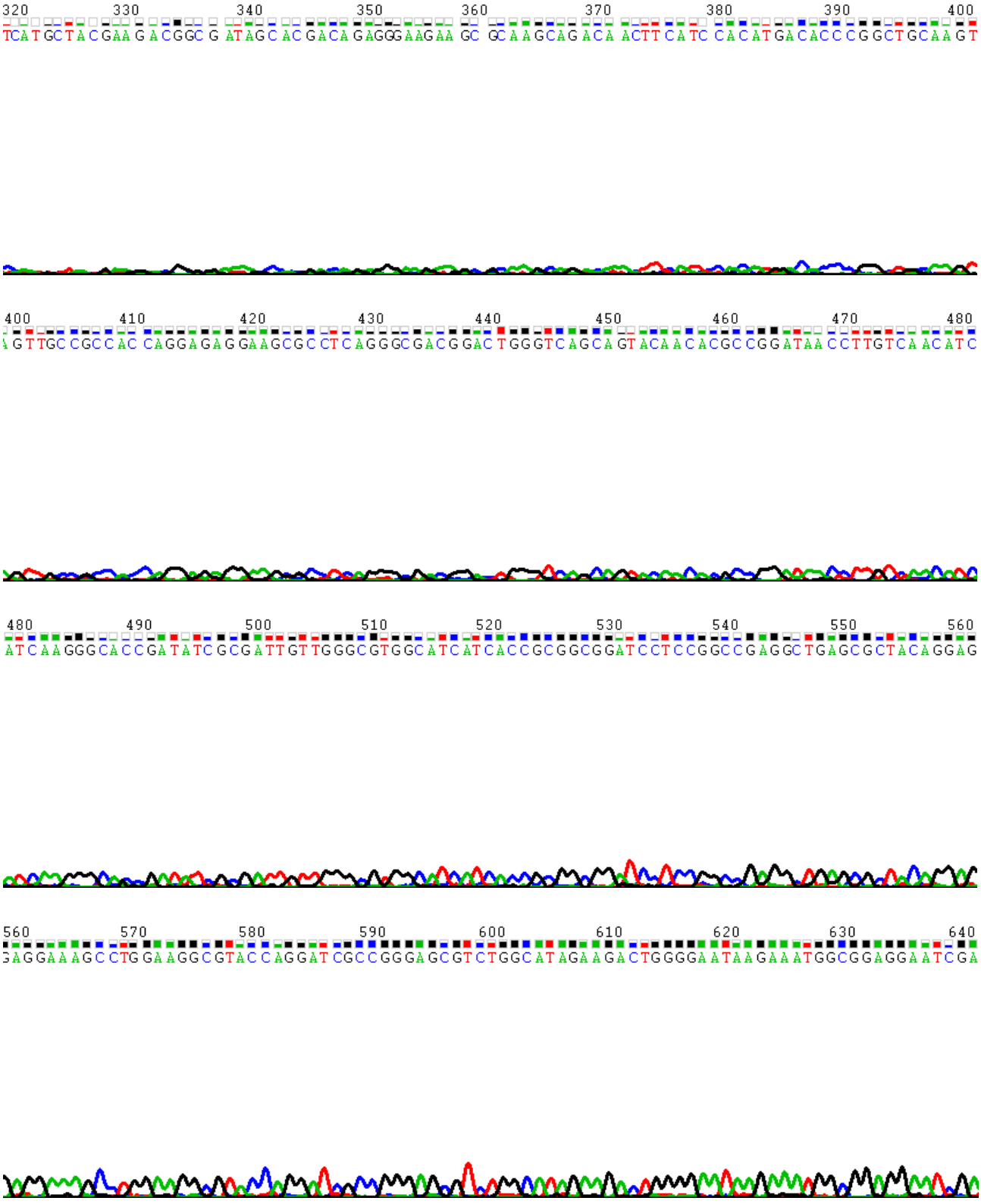
	Vitamin				Carbon source / stress treatment					
Progeny	**paba pyro**	**paba**	**pyro**	**Neither vitamin**	**Glycerol**	**Ethanol**	**Galacturonic acid**	**0.5 M NaCl**	**UV Irradiation (6 mins)**	**50 mM Arginine (N-source)**
GO281	G	G	X	X	P	P	P	P	P	P
WP1	G	X	G	G	G	G	G	G	G	G
51	G	X	G	X	G	G	G	G	G	G
52	G	X	G	X	G	G	G	G	G	G
53	G	X	G	X	G	G	G	G	G	G
54	G	X	G	X	G	G	G	G	G	G
55	G	X	G	X	P	P	P	G	G	G
56	G	G	X	X	G	G	G	G	G	G
57	G	X	G	X	G	G	G	G	G	G
58	G	X	G	X	G	G	G	G	G	G
59	G	X	G	X	G	G	G	G	G	G
60	G	G	X	X	P	P	P	P	P	P
61	G	G	G	G	G	G	G	G	G	G
62	G	G	X	X	P	P	P	P	P	P
63	G	X	G	X	G	G	G	G	G	G
64	G	X	G	X	G	G	G	G	G	G
65	G	X	G	X	G	G	G	G	G	G
66	G	G	G	G	P	P	P	G	G	G
67	G	X	X	X	G	G	G	G	G	G
68	G	X	G	X	G	G	G	G	G	G
69	G	X	X	X	P	P	P	G	G	G
60	G	G	X	X	P	P	P	G	G	G
71	G	X	X	X	G	G	G	G	G	G
72	G	X	G	X	P	P	P	P	P	P
73	G	G	X	X	P	P	P	G	G	G
74	G	X	G	X	P	P	P	G	G	G
75	G	X	G	X	G	G	G	G	G	G
76	G	G	G	G	G	G	G	G	G	G
77	G	X	G	X	G	G	G	G	G	G
78	G	X	X	X	P	P	P	P	P	P
79	G	G	G	G	G	G	G	G	G	G
70	G	G	G	G	P	P	P	G	G	G
81	G	G	X	X	P	P	P	P	P	P
82	G	G	X	X	G	G	G	P	P	P
83	G	X	G	X	G	G	G	G	G	G
84	G	X	X	X	P	P	P	G	G	G
85	G	G	X	X	G	G	G	P	P	P
86	G	X	X	X	P	P	P	P	P	P
87	G	G	G	G	G	G	G	P	P	P
88	G	G	G	G	G	G	G	G	G	G
89	G	G	G	G	G	G	G	P	P	P
90	G	X	X	X	P	P	P	P	P	P
91	G	G	X	X	G	G	G	P	P	P
92	G	G	X	X	P	P	P	P	P	P
93	G	G	X	X	G	G	G	G	G	G
94	G	G	X	X	P	P	P	G	G	G
95	G	G	G	G	G	G	G	G	G	G
96	G	G	X	X	G	G	G	G	G	G
97	G	G	G	G	P	P	P	G	G	G
98	G	G	G	G	P	P	P	G	G	G
99	G	G	X	X	G	G	G	P	P	P
100	G	G	X	X	G	G	G	P	P	P

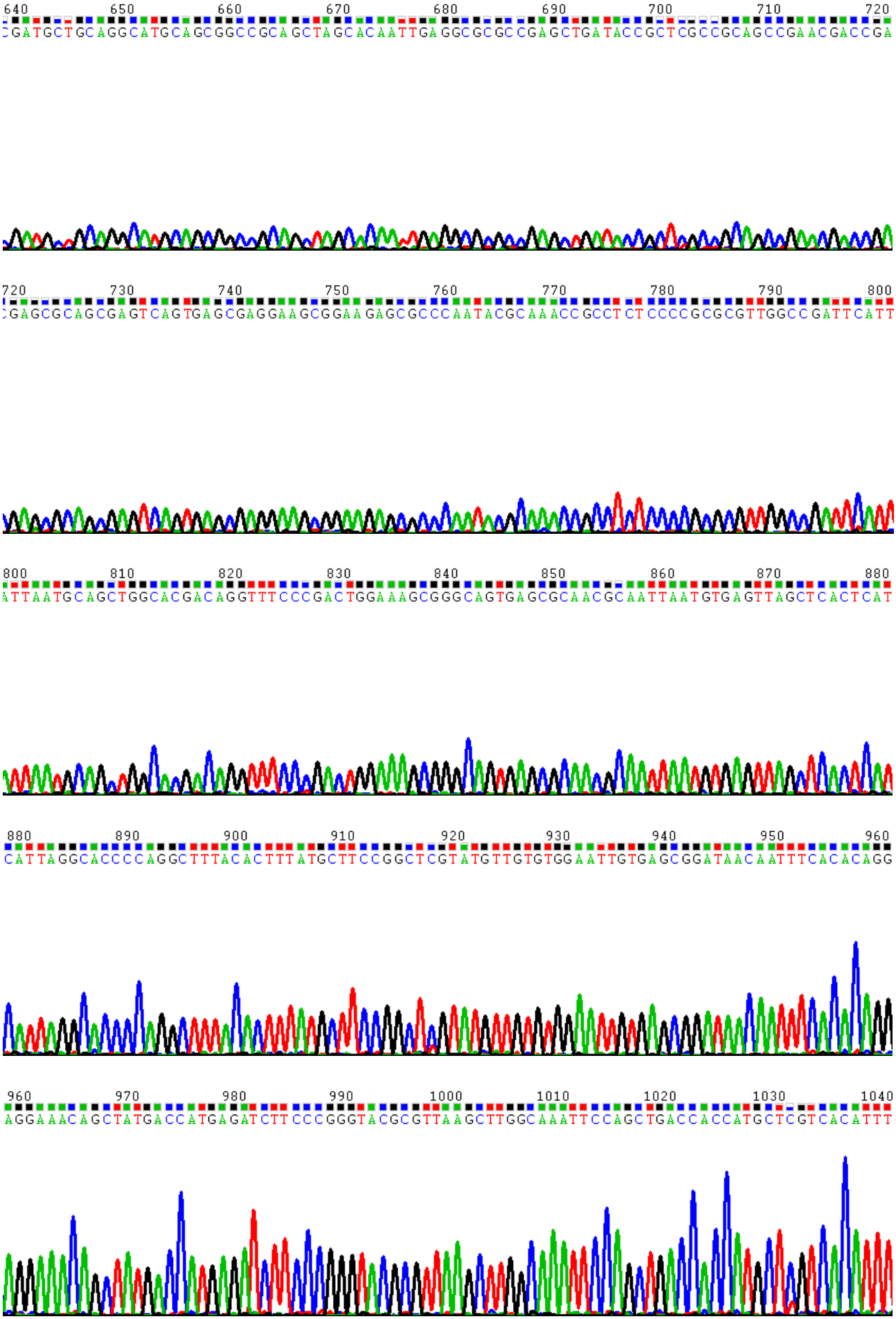
Appendix 4 – Restriction map of pDJB3. The plasmid was constructed by Ballance and Turner (1985) and has a total size of 11,050 bp. The smallest divisions on the map represent 200 bp and larger divisions represent 1 kb. The ampicillin resistance genes (AMP) allow selection of pDJB3-transformed *E. coli* JM109, a strain that has no natural resistance to ampicillin. The *pyr4* gene, from *N. crassa*, can complement the defective *pyrG* gene in *A. nidulans* strains harbouring the *pyrG89* mutation by encoding orotidine-5'-phosphate decarboxylase. This enzyme is involved in the pyrimidine biosynthetic pathway and restores wild-type growth of *pyrG89* strains in the absence of uridine and uracil. The 3.5 kb *ans1* sequence, which is reiterated in the *A. nidulans* genome, is known to dramatically increase the transformation efficiency of the plasmid due to homologous recombination events.

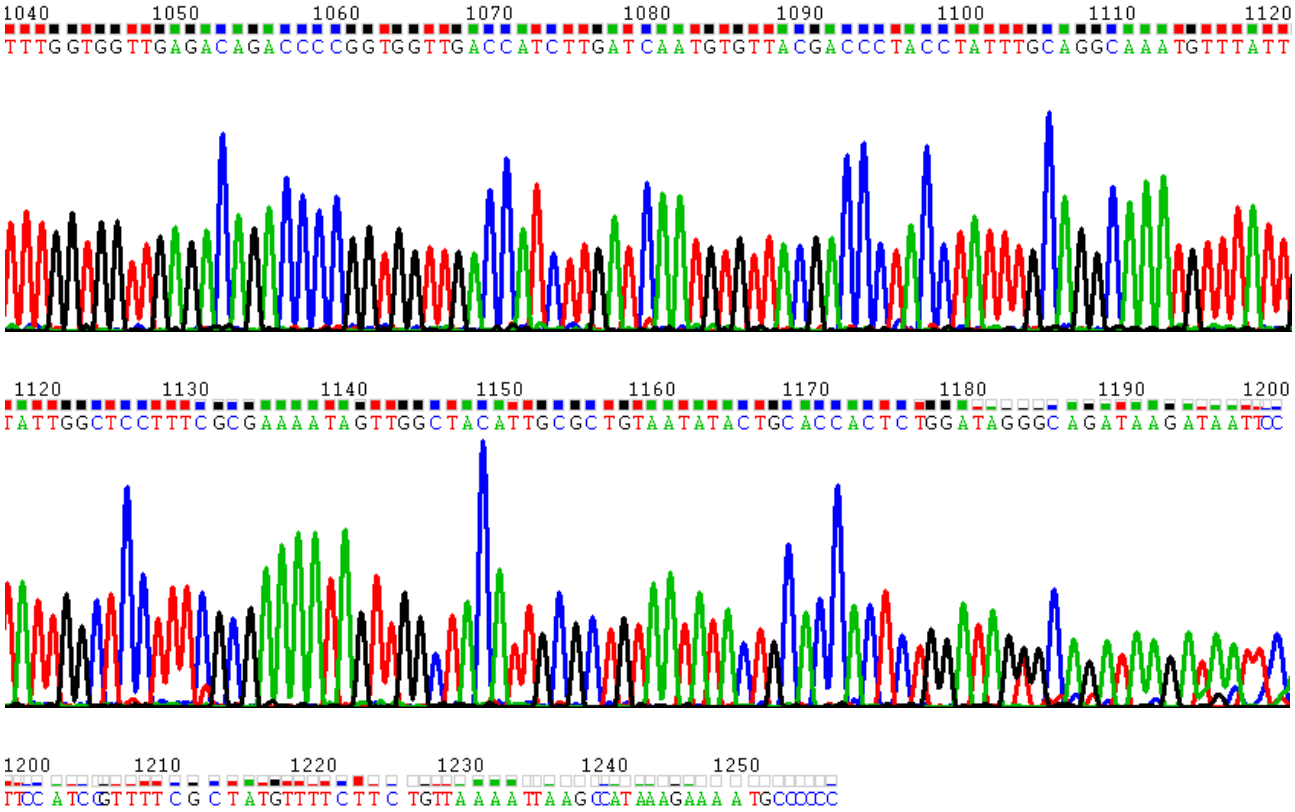


Appendix 5 – Sequence data for regions A and B of the PCR assay product from A2.

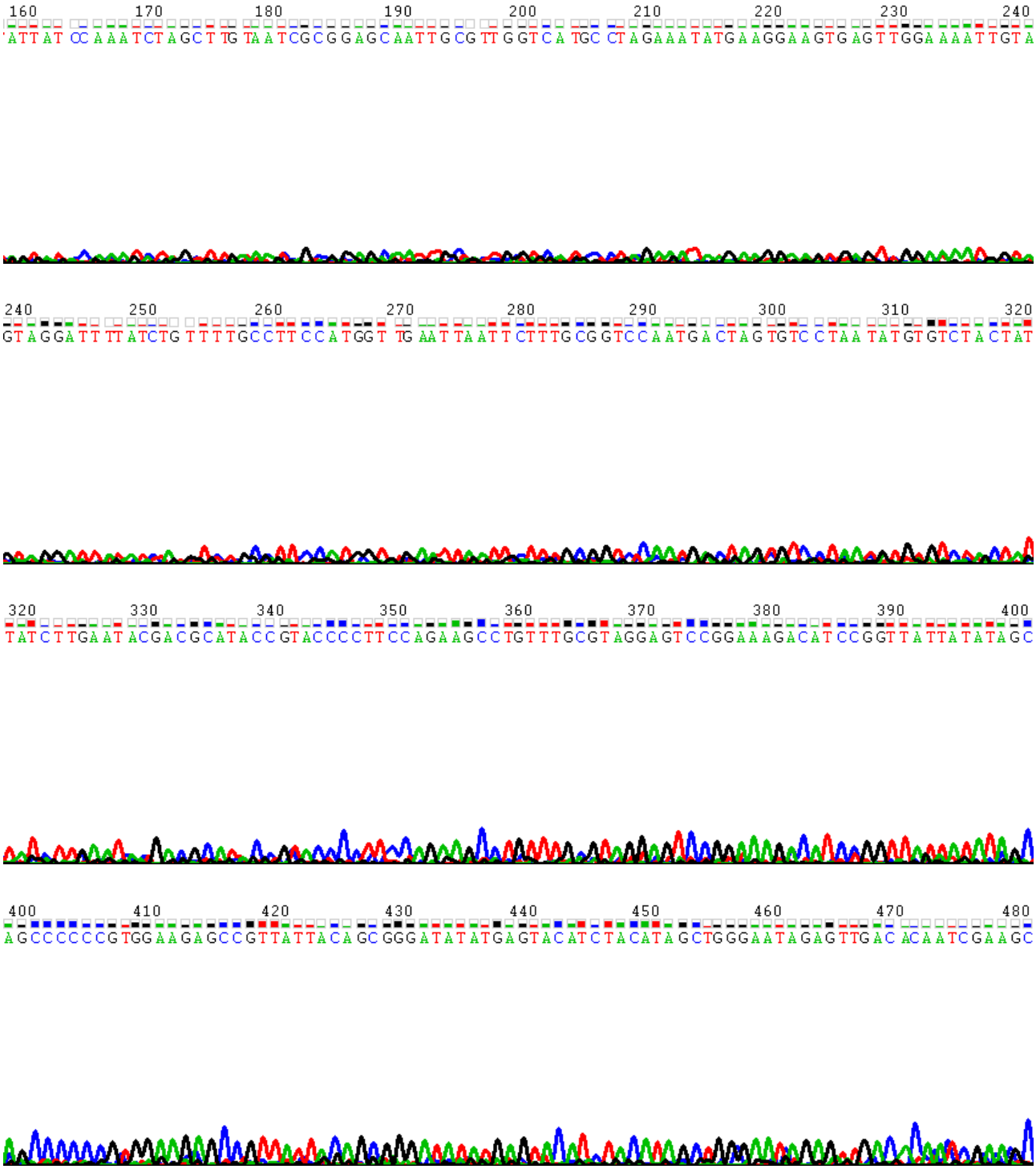
Sequence data for region A. Direct sequencing was carried out by the Dundee Sequencing Centre (http://www.dnaseq.co.uk) using the primer designated SA. The sequence shown is the reverse complement of the data obtained. The first 320 nucleotides are omitted due to low signal.

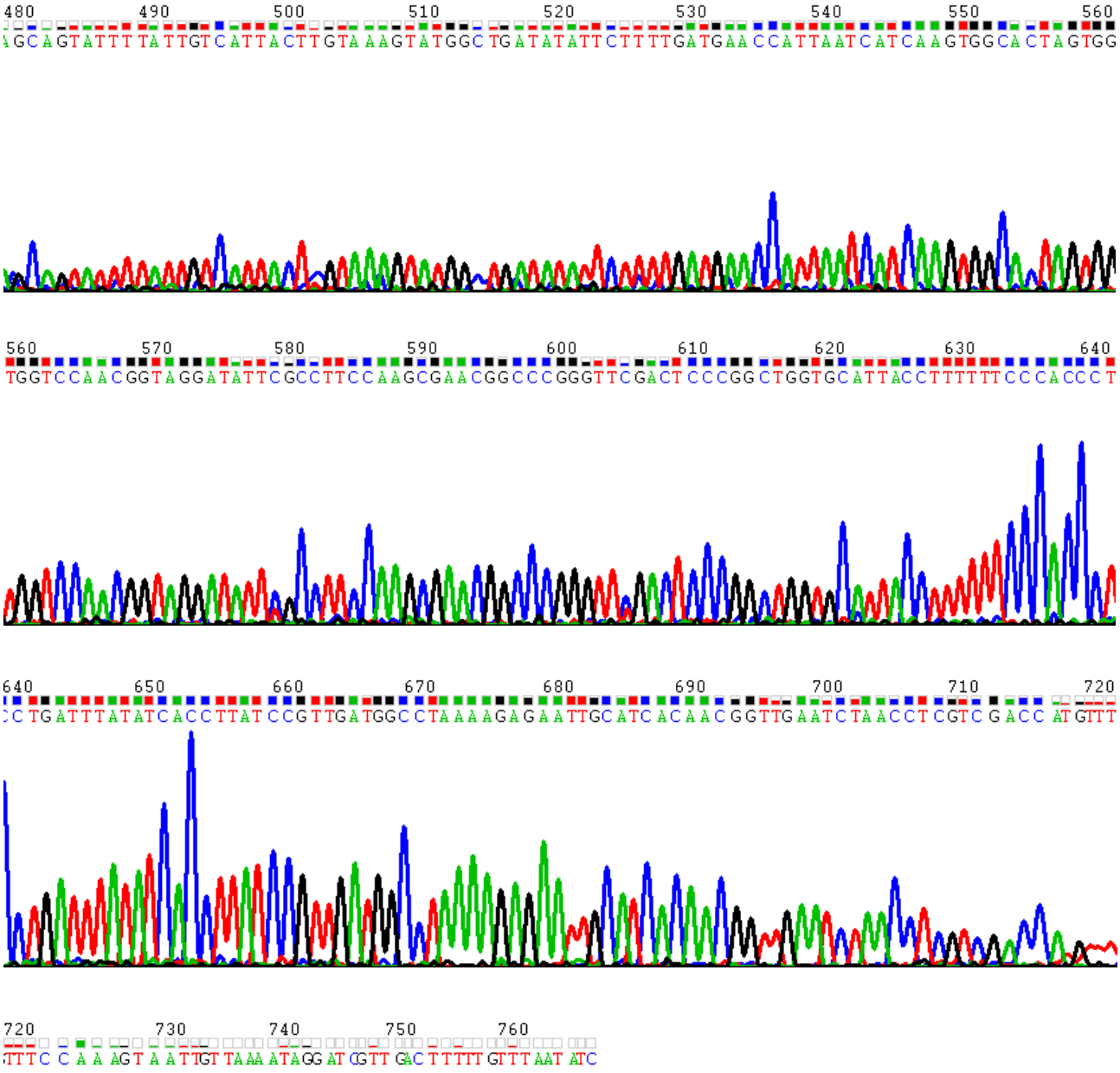



640 650 660 670 680 690 700 710 720
CGATGCTGCAGGCATGCAGCGGCCGCAGCTAGCACAATTGAGGCGCGCCGAGCTGATACCGCTCGCCGCAGCCGAACGACCGA
720 730 740 750 760 770 780 790 800
CGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATT
800 810 820 830 840 850 860 870 880
ATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCAT
880 890 900 910 920 930 940 950 960
CATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGG
960 970 980 990 1000 1010 1020 1030 1040
AGGAAACAGCTATGACCATGAGATCTTCCCGGGTACGCGTTAAGCTTGGCAAATTCCAGCTGACCACCATGCTCGTCACATTT


1040 1050 1060 1070 1080 1090 1100 1110 1120
TTTGGTGGTTGAGACAGACCCCGGTGGTTGACCATCTTGATCAATGTGTTACGACCCTACCTATTTGCAGGCAAATGTTTATT
1120 1130 1140 1150 1160 1170 1180 1190 1200
TATTGGCTCCTTTCGCGAAAAATAGTTGGCTACATTGCGCTGTAATATACTGCACCACTCTGGATAGGGCAGATAAGATAATTCC
1200 1210 1220 1230 1240 1250
TTCC ATCCGTTTTCGCTATGTTTTCTTC TGTTAAAATTAAGCCATAAAGAAAA ATGCCCCCC

Sequence data for region B. Direct sequencing was carried out by the Dundee Sequencing Centre (http://www.dnaseq.co.uk) using the primer designated SB. The sequence shown is the reverse complement of the data obtained. The first 160 nucleotides are omitted due to low signal.

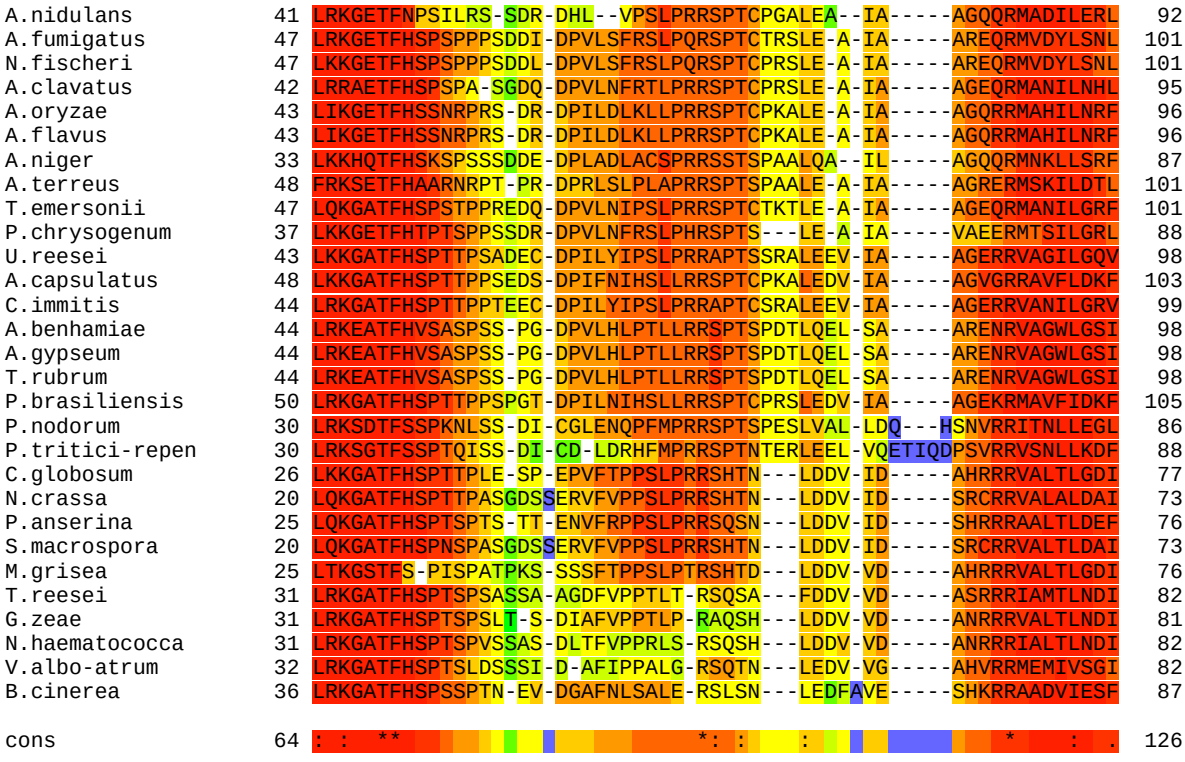


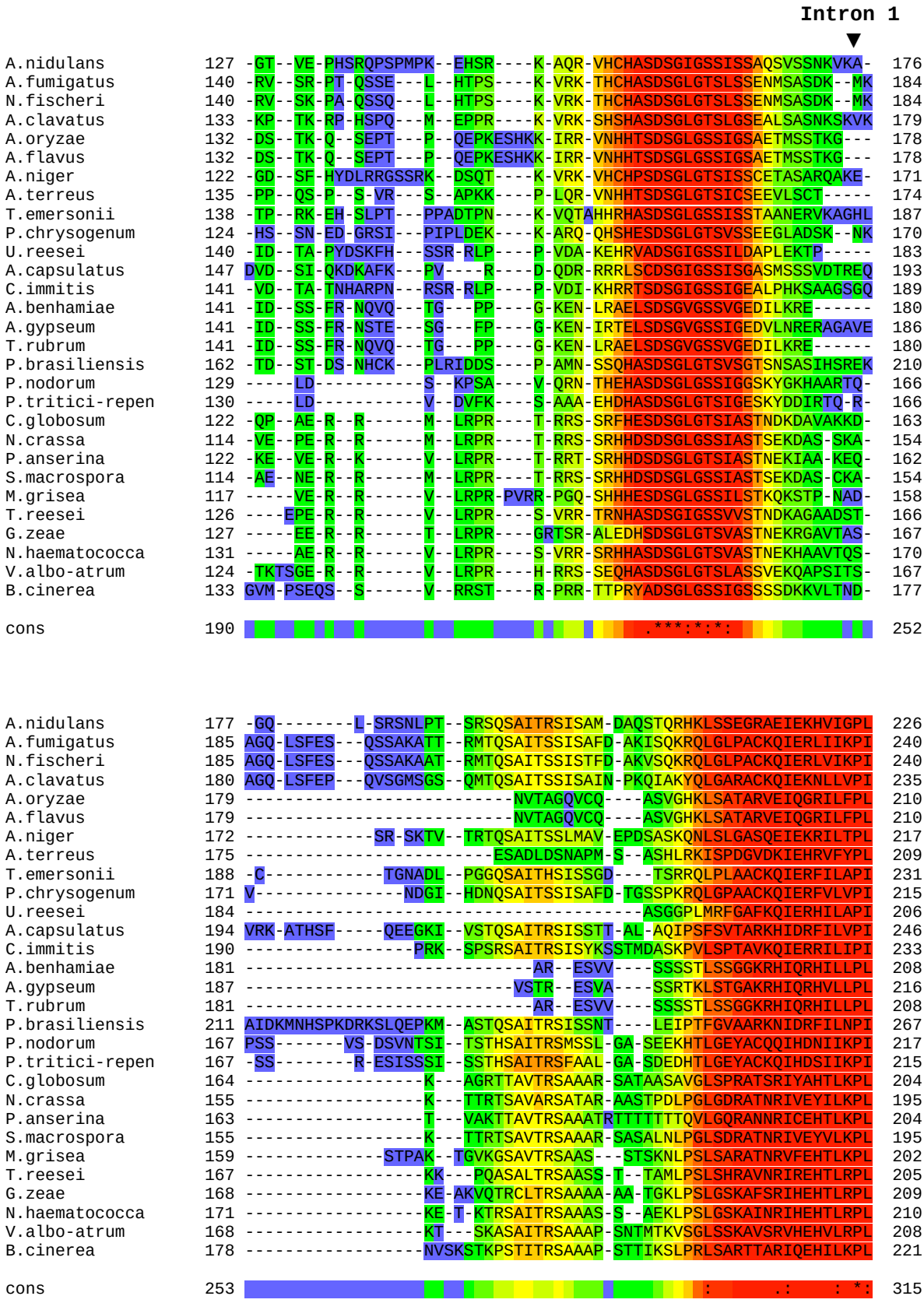

480 490 500 510 520 530 540 550 560
AGCAGTATTTTATTGTCATTACTTGTAAAGTATGGCTGATATATTCTTTTGATGAACCATTAATCATCAAGTGGCACTAGTGG
560 570 580 590 600 610 620 630 640
TGGTCCAACGGTAGGATATTCGCCTTCCAAGCGAACGGCCCGGGTTCGACTCCCGGCTGGTGCATTACCTTTTTTCCCACCCT
640 650 660 670 680 690 700 710 720
CCTGATTTATATCACCTTATCCGTTGATGGCCTAAAAAGAGAATTGCATCACAACGGTTGAATCTAACCTCGTCGACCATGTTT
720 730 740 750 760
TTTCCCAAAGTAATTGTTAAAATAGGATCGTTGACTTTTTGTTTAATATC

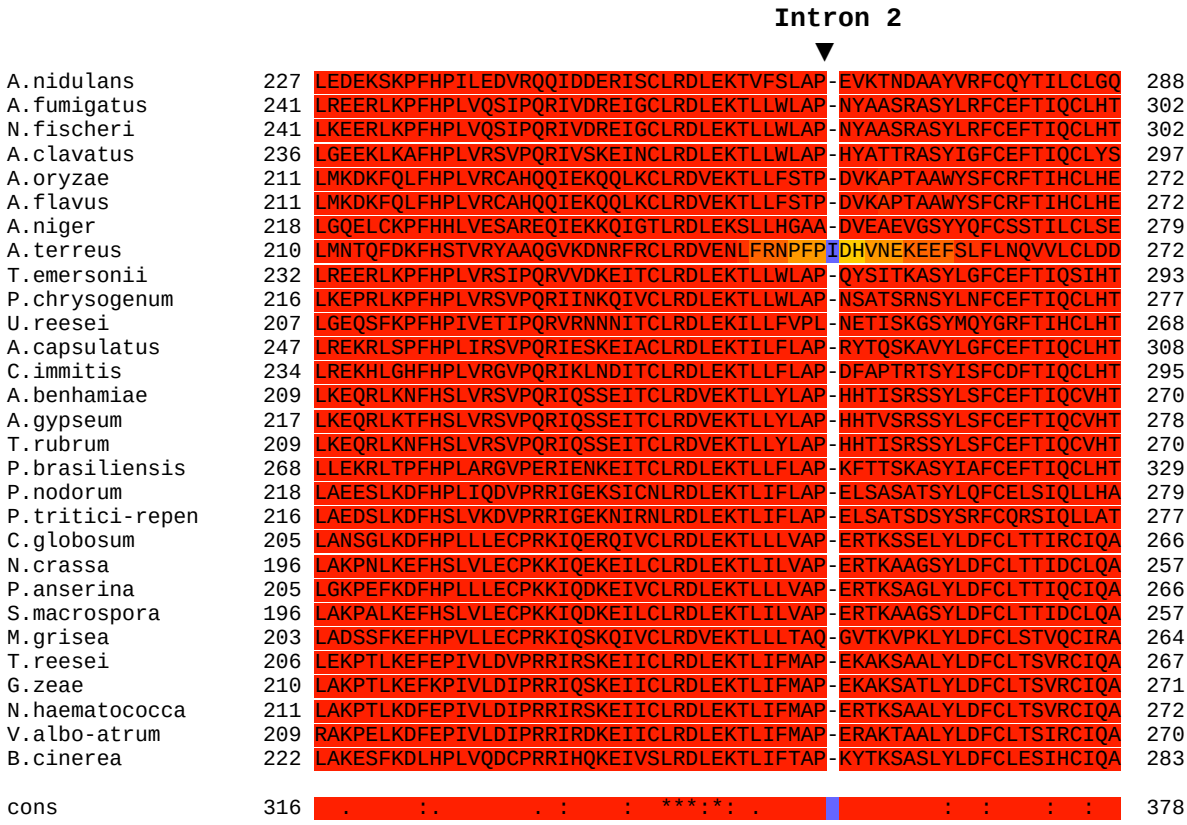
Appendix 6 – Alignment of 29 manually annotated StzA proteins using Advanced TCoffee

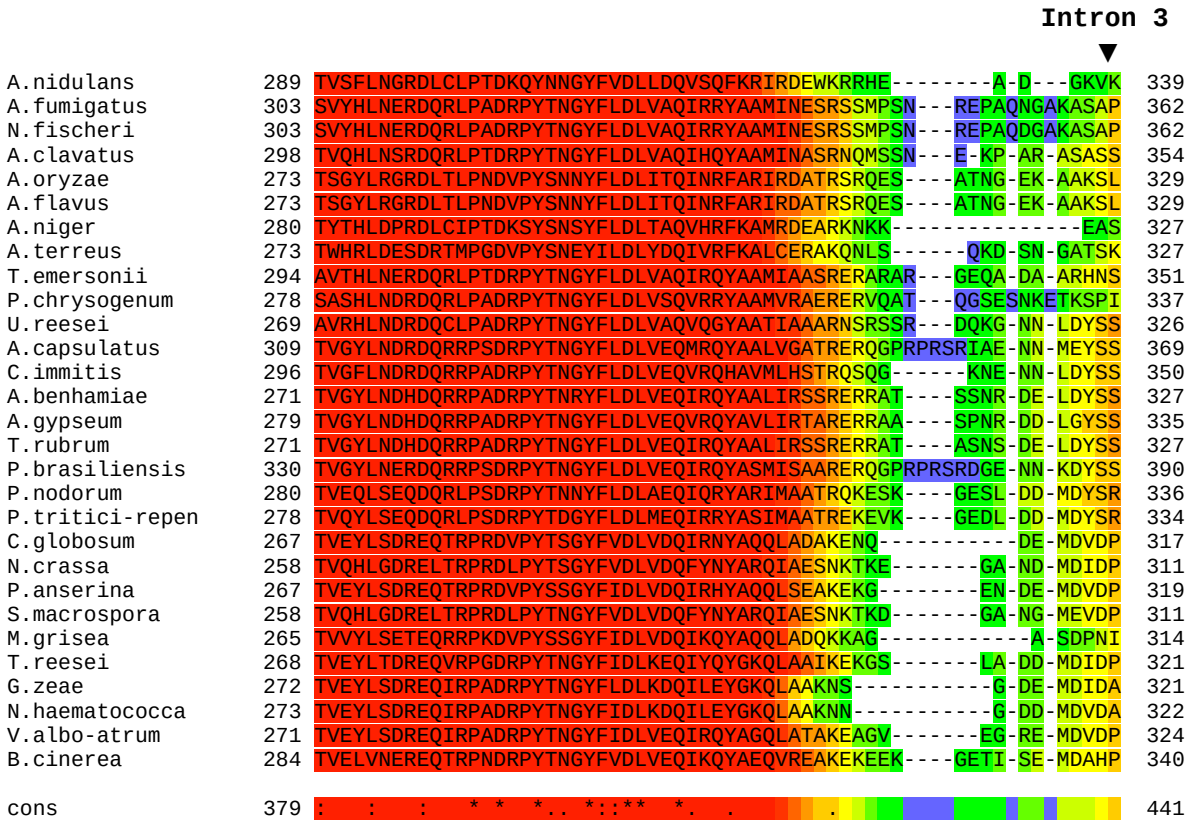
http://tcoffee.vital-it.ch/cgi-bin/Tcoffee/tcoffee_cgi/index.cgi

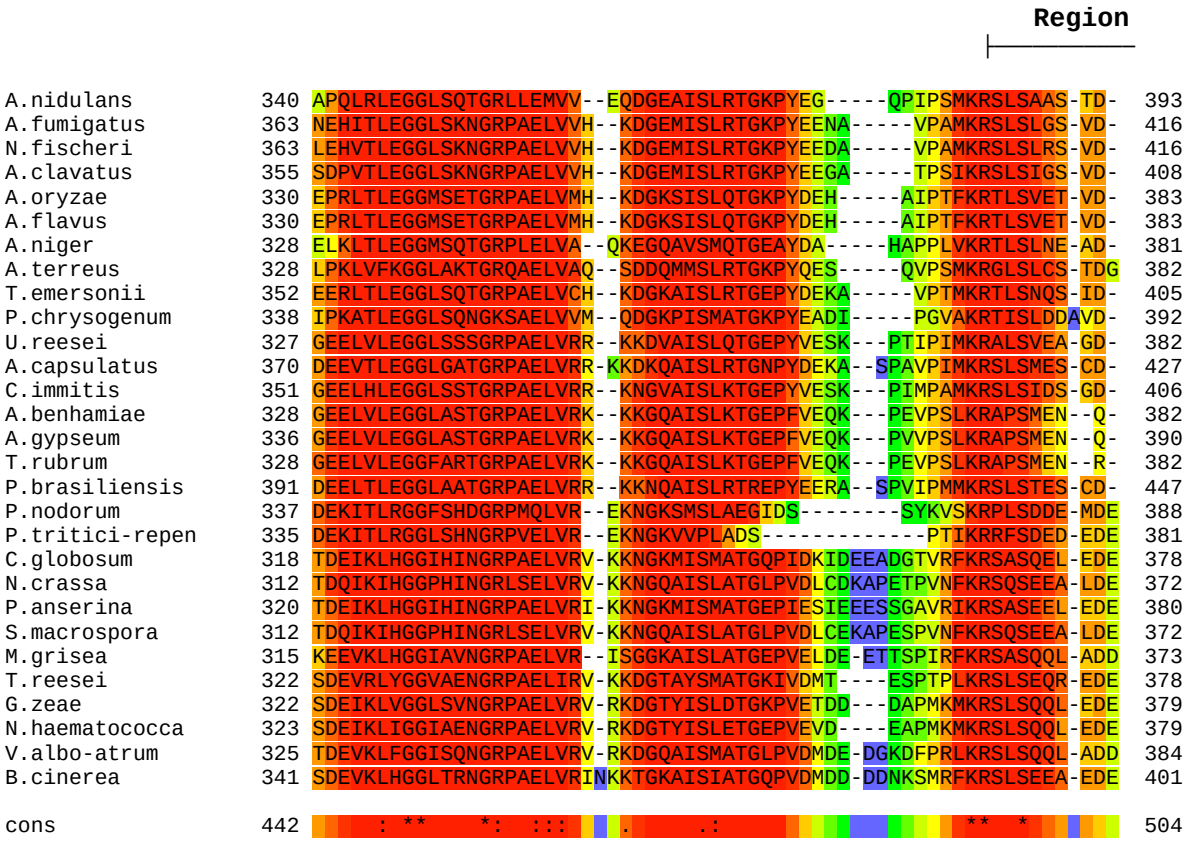
```
T-COFFEE, Version_9.02.r1228 (2012-02-16 18:15:12 - Revision 1228 - Build 336)
Cedric Notredame
CPU TIME:146 sec.
SCORE=82
*
 BAD AVG GOOD
*
A.nidulans              :  81
A.fumigatus             :  82
N.fischeri              :  82
A.clavatus              :  83
A.oryzae                :  83
A.flavus                :  83
A.niger                 :  83
A.terreus               :  81
T.emersonii             :  83
P.chrysogenum           :  81
U.reesei                :  84
A.capsulatus            :  81
C.immitis               :  82
A.benhamiae             :  83
A.gypseum               :  82
T.rubrum                :  82
P.brasiliensis          :  81
P.nodorum               :  82
P.tritici-repen         :  82
C.globosum              :  84
N.crassa                :  83
P.anserina              :  82
S.macrospora            :  83
M.grisea                :  82
T.reesei                :  82
G.zeae                  :  82
N.haematococca          :  82
V.albo-atrum            :  83
B.cinerea               :  81
cons                    :  82


A.nidulans              1 MSPAQD---SESIKAHPRRRP-FRA-A-R-PSLV--PA-----------EEQS-PSL-PPLR   40
A.fumigatus             1 MSTSQA----EHTRVHPRRRP-LKT--A-PPSLSA-STSIESYVA------DSS-SSS-PNLS   46
N.fischeri              1 MSTSQA----EHTRVHPRRRP-LKT--A-PPSLSA-STSLESYVA------DSS-SSS-PSLS   46
A.clavatus              1 MSAPQA----EPTRIHPRRRP-LKT--A-PPSLIT-SASPE-----------SH-DGP-SSPY   41
A.oryzae                1 MSSAQPV--DESDRRHPRRRV-LNR-P--PPSLSS-DSS------------PPH-LLP-SNLR   42
A.flavus                1 MSSAQPV--DESDRRHPRRRV-LNR-P--PPSLSS-DSS------------PPH-LLP-SNLR   42
A.niger                 1 MKS-----------QNPRRGR-HVA-R-P-PSLA--ST-----------ADVH-ASR-SRYQ   32
A.terreus               1 MSSPKDSQYDTTLRRHPRRRA-ALS-TS-PPSLLN-SSNCR----------SVD-STP-SNPL   47
T.emersonii             1 MSTQTQ-----QSQVHPRRRP-YKA--G-IPSLST-SSASDQQDAK-----SSS-PCP-PTPR   46
P.chrysogenum           1 MSAIQT----EPSRVHPRRRP-VLN-TA-LPTLSS-D------------------LPD-SKMP   36
U.reesei                1 MSEP---------QSHPRRRP-FNG-AA-LPSLLT-SSNHGNGPA------SHV-VAP-YKRM   42
A.capsulatus            1 MSAP---------QVHSRRRP-LQH-KN-LPHLSI-TSSSSTSDQSSSDS-TCR-LTP-HSMK   47
C.immitis               1 MSES---------QSHPRRRP-FNK-AAPLPPLLT-SSDHGNGSS------NRA-VRP-HNMK   43
A.benhamiae             1 MSSDL--------SISPRRRP-FLN-AN-LPRLST-GHCDGRSPS------NSS-NTV-RHAT   43
A.gypseum               1 MSSDL--------SISPRRRP-FLN-AN-LPRLST-GHCDGRSPS------NSS-NTV-RHAT   43
T.rubrum                1 MSSDL--------SISPRRRP-FLN-AN-LPRLST-SHCDGRSPS------NSS-NTV-RHAT   43
P.brasiliensis          1 MSAP---------RVHPRRRP-IYN-SAPIPRLPT-SSSPLSSSSDLSDPACRH-SAL-HSMQ   49
P.nodorum               1 MSAPHG------QNSHPRRRP-ING-Q-T----------------------SPL-N-T-HGMS   29
P.tritici-repen         1 MSAPHG------QNCHPRRRP-LKG-T-A----------------------SPL-S-I-HGMQ   29
C.globosum              1 MSS-----------NPRRTP-VTR-S-G-S--------------------RGL-SIK-TNVV   25
N.crassa                1 MS------------NPRRRP-AGG-------------------------L-TLK-TNML   19
P.anserina              1 MSS-----------NPRRTP-MTR-P-D-S--------------------RGL-SLK--TNT   24
S.macrospora            1 MS------------NPRRRP-AGG--------------------------L-TLK-TNVL   19
M.grisea                1 MS------------NPRRTP-MTR-P-D-S----------------------VL-PLKTTLAV   24
T.reesei                1 MSFS----------NPRRRTPVTR-P-G-T-----DC-----------EHGL-SLK-TTMT   30
G.zeae                  1 MSFS----------HPRRRTPVTR-P-D-C-----DT-----------ENAL-SFK-NSST   30
N.haematococca          1 MSFS----------HPRRRTPVTR-P-G-S-----DT-----------ENAL-SLK-TNST   30
V.albo-atrum            1 MSCQ----------NPRRRSPVTR-V-G-D-----AS-----------SNGLTSLK-TNMT   31
B.cinerea               1 MSFT----------HPRRTT-KTTSI-A-VTLLANSS-----------SRPT-VNT-TMAS   35

cons                    1 *.  .**                                                          63
```

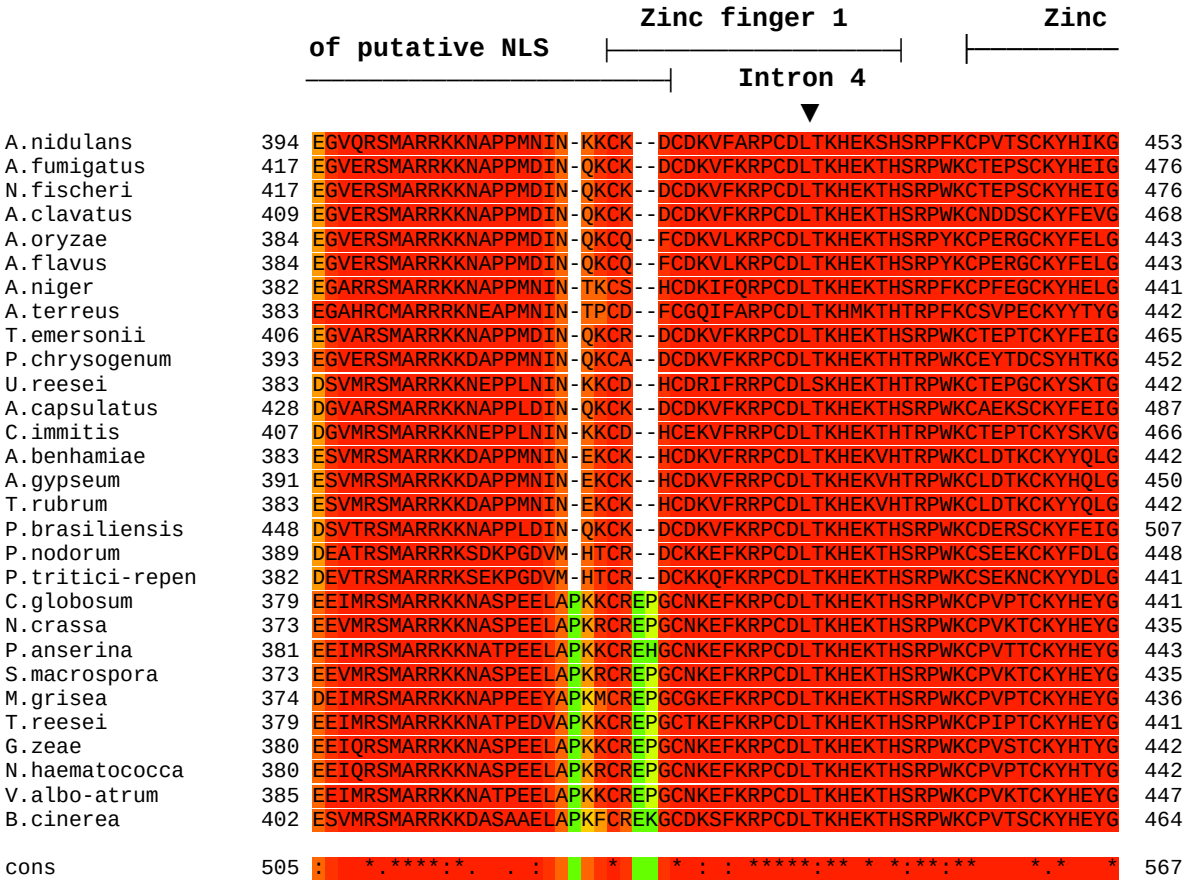

A.nidulans 41 LRKGETFNPSILRS-SDR-DHL--VPSLPRRSPTCPGALEA--IA-----AGQQRMADILERL 92
A.fumigatus 47 LRKGETFHSPSPPPSDDI-DPVLSFRSLPQRSPTCTRSLE-A-IA-----AREQRMVDYLSNL 101
N.fischeri 47 LKKGETFHSPSPPPSDDL-DPVLSFRSLPQRSPTCPRSLE-A-IA-----AREQRMVDYLSNL 101
A.clavatus 42 LRRAETFHSPSPA-SGDQ-DPVLNFRTLPRRSPTCPRSLE-A-IA-----AGEQRMANILNHL 95
A.oryzae 43 LIKGETFHSSNRPRS-DR-DPILDLKLLPRRSPTCPKALE-A-IA-----AGQRRMAHILNRF 96
A.flavus 43 LIKGETFHSSNRPRS-DR-DPILDLKLLPRRSPTCPKALE-A-IA-----AGQRRMAHILNRF 96
A.niger 33 LKKHQTFHSKSPSSSDDE-DPLADLACSPRRSSTSPAALQA--IL-----AGQQRMNKLLSRF 87
A.terreus 48 FRKSETFHAARNRPT-PR-DPRLSLPLAPRRSPTSPAALE-A-IA-----AGRERMSKILDTL 101
T.emersonii 47 LQKGATFHSPSTPPREDQ-DPVLNIPSLPRRSPTCTKTLE-A-IA-----AGEQRMANILGRF 101
P.chrysogenum 37 LKKGETFHTPTSPPSSDR-DPVLNFRSLPHRSPTS---LE-A-IA-----VAEERMTSILGRL 88
U.reesei 43 LKKGATFHSPTTPSADEC-DPILYIPSLPRRAPTSSRALEEV-IA-----AGERRVAGILGQV 98
A.capsulatus 48 LKKGATFHSPTTPPSEDS-DPIFNIHSLLRRSPTCPKALEDV-IA-----AGVGRRAVFLDKF 103
C.immitis 44 LRKGATFHSPTTPPTEEC-DPILYIPSLPRRAPTCSRALEEV-IA-----AGERRVANILGRV 99
A.benhamiae 44 LRKEATFHVSASPSS-PG-DPVLHLPTLLRRSPTSPDTLQEL-SA-----ARENRVAGWLGSI 98
A.gypseum 44 LRKEATFHVSASPSS-PG-DPVLHLPTLLRRSPTSPDTLQEL-SA-----ARENRVAGWLGSI 98
T.rubrum 44 LRKEATFHVSASPSS-PG-DPVLHLPTLLRRSPTSPDTLQEL-SA-----ARENRVAGWLGSI 98
P.brasiliensis 50 LRKGATFHSPTTPPSPGT-DPILNIHSLLRRSPTCPRSLEDV-IA-----AGEKRMAVFIDKF 105
P.nodorum 30 LRKSDTFSSPKNLSS-DI-CGLENQPFMPRRSPTSPESLVAL-LDQ---HSNVRRITNLLEGL 86
P.tritici-repen 30 LRKSGTFSSPTQISS-DI-CD-LDRHFMPRRSPTNTERLEEL-VQETIQDPSVRRVSNLLKDF 88
C.globosum 26 LKKGATFHSPTTPLE-SP-EPVFTPPSLPRRSHTN---LDDV-ID-----AHRRRVALTLGDI 77
N.crassa 20 LQKGATFHSPTTPASGDSSERVFVPPSLPRRSHTN---LDDV-ID-----SRCRRVALALDAI 73
P.anserina 25 LQKGATFHSPTSPTS-TT-ENVFRPPSLPRRSQSN---LDDV-ID-----SHRRRAALTLDEF 76
S.macrospora 20 LQKGATFHSPNSPASGDSSERVFVPPSLPRRSHTN---LDDV-ID-----SRCRRVALTLDAI 73
M.grisea 25 LTKGSTFS-PISPATPKS-SSSFTPPSLPTRSHTD---LDDV-VD-----AHRRRVALTLGDI 76
T.reesei 31 LRKGATFHSPTSPSASSA-AGDFVPPTLT-RSQSA---FDDV-VD-----ASRRRIAMTLNDI 82
G.zeae 31 LRKGATFHSPTSPSLT-S-DIAFVPPTLP-RAQSH---LDDV-VD-----ANRRRVALTLNDI 81
N.haematococca 31 LRKGATFHSPTSPVSSAS-DLTFVPPRLS-RSQSH---LDDV-VD-----ANRRRIALTLNDI 82
V.albo-atrum 32 LRKGATFHSPTSLDSSSI-D-AFIPPALG-RSQTN---LEDV-VG-----AHVRRMEMIVSGI 82
B.cinerea 36 LRKGATFHSPSSPTN-EV-DGAFNLSALE-RSLSN---LEDFAVE-----SHKRRAADVIESF 87
cons 64 126
A.nidulans 93 DLNS-------G-----------TT--S-TSDE--NDDLPVPKGLLRLHLQTQARRE----- 126
A.fumigatus 102 NLNSLE--PSS-----------PL---SDKDN--GDDLPVPRAILQAHIDSQSMAD----- 139
N.fischeri 102 NLNSLE--PSS-----------PL---SDKDN--GDDLPVPRAILQAHIDSQSMAD----- 139
A.clavatus 96 DLDSLD--ATG-----------PL---N-SSS--DNDLPVPRGVLQAHFDSEPLAD----- 132
A.oryzae 97 DLDSLS-----------------TR---DSLES--QDELPVPRGILRTHVKSSASKE----- 131
A.flavus 97 DLDSLS-----------------TR---DSLES--QDELPVPRGILRTHVKSSASKE----- 131
A.niger 88 DLDS-------S-----------DS--A-PRKE--HDGLPVPRGYLEIHFERSQSGK----- 121
A.terreus 102 DLDTFT-----------------P----SESV--DEELPVPRSVLQLHFDSLHISD----- 134
T.emersonii 102 DFDSLS--PS-----------KA---TNRQE--EDDLPVPRGILQAHVGQQ-RMQ----- 137
P.chrysogenum 89 TLEPTQ--DQS-----------S------EVA--GGESPLDSGLGRNNSKSMPDSP----- 123
U.reesei 99 ERDLAGIDGKST----------RN---RPSFP--GDDFPVPRGLLHARVADTDPMD----- 139
A.capsulatus 104 ERSPYGLGSDLS----------APRL-DNTLC--DNDSPVPRGMPEARMTNTDPMN----- 146
C.immitis 100 ERDLAGRDGEPS----------RS---RPSFP--GDDSPVPRGILAAHIANTDAMD----- 140
A.benhamiae 99 ERNSTGPGNKSL----------SK---QSSLEDDKRNLKALRGLSK-RLTSANLME----- 140
A.gypseum 99 ERNSTGPGNKSL----------SK---QSSLEDVNSSLKAPRGLNT-RLTSANLME----- 140
T.rubrum 99 ERNSTGLGNKSL----------SK---QSSLEDDKRNLKALRGLSK-RLTSANLME----- 140
P.brasiliensis 106 ERNLSGLGTCSPSTASSSSSSSSSPRNRHSILD--DDDLPVPRAILDAHVIHADTMD----- 161
P.nodorum 87 DKQLSGHK-AAS----------AIA--ANILAD--PEVLPVPSFVLDNATLDTTPME----- 128
P.tritici-repen 89 EARVAGHK-TSN----------T-G--ASILSD--PDVFPVPSLFLDNTSMDCTPME----- 129
C.globosum 78 DKTLAGLRNDDA----------PLSP-GRKTLR--DEILPLPRGVLDHTLDSVMV-K----E 121
N.crassa 74 ERQLASS--NDT----------FAS--A--SRS--DKCIPPPRGLLERNLDSPIMPK----E 113
P.anserina 77 DRTLAGLSISDS----------PSSAAARKILR--EDSPPIPRGILNHTLDTVMA-K----G 121
S.macrospora 74 TRQLASS--NDT----------FVS--A--KQS--GECIPPPRGLLQRNLDSPIMPQ----E 113
M.grisea 77 EKTLAGMS-LDS----------PSA--N-KAFR--DNSYPLPRGLLDPTLIGKEM------D 116
T.reesei 83 DEALSKASLSDK----------SP--RPKPLR--DTSLPVPRGFLEPPVVDPAMNK----Q 125
G.zeae 82 DEALAKTQELSLSS----------TS--KPMTLR--DTGLPIPRGFLEGPIVDPKMTK----E 126
N.haematococca 83 DEALAKTEELSLSP-------RSPKS--KPKSLR--DTSLPIPRGFLEGPIVDPAMAK----E 130
V.albo-atrum 83 ETSLNLND--TP----------RP--ASKPSR--DECLPRTNGFLGRPTVDPAMAK----D 123
B.cinerea 88 ELTVAGN--QSP----------ST--PRRRFR--DESDPVPRGVVNVENPSTKAYNTELMD 132
cons 127 189

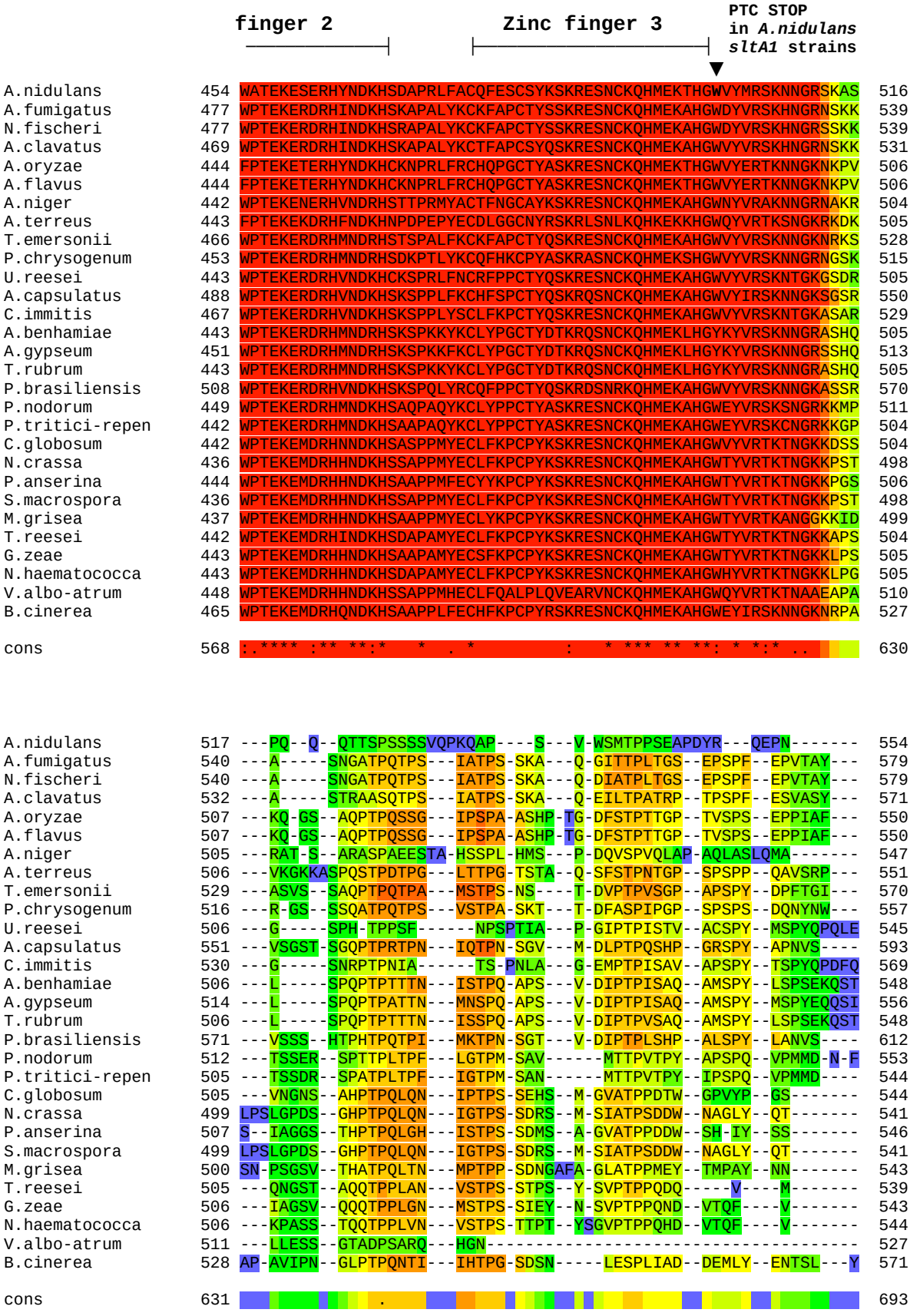

Intron 1
A.nidulans 127 -GT--VE-PHSRQPSPMPK--EHSR----K-AQR-VHCHASDSGIGSSISSAQSVSSNKVKA- 176
A.fumigatus 140 -RV--SR-PT-QSSE---L--HTPS----K-VRK-THCHASDSGLGTSLSSENMSASDK--MK 184
N.fischeri 140 -RV--SK-PA-QSSQ---L--HTPS----K-VRK-THCHASDSGLGTSLSSENMSASDK--MK 184
A.clavatus 133 -KP--TK-RP-HSPQ---M--EPPR----K-VRK-SHSHASDSGLGTSLGSEALSASNKSKVK 179
A.oryzae 132 -DS--TK-Q--SEPT---P--QEPKESHKK-IRR-VNHHTSDSGLGSSIGSAETMSSTKG--- 178
A.flavus 132 -DS--TK-Q--SEPT---P--QEPKESHKK-IRR-VNHHTSDSGLGSSIGSAETMSSTKG--- 178
A.niger 122 -GD--SF-HYDLRRGSSRK--DSQT----K-VRK-VHCHPSDSGLGTSISSCETASARQAKE- 171
A.terreus 135 -PP--QS-P--S-VR---S--APKK----P-LQR-VNHHTSDSGLGTSICSEEVLSCT----- 174
T.emersonii 138 -TP--RK-EH-SLPT---PPADTPN----K-VQTAHHRHASDSGLGSSISSTAANERVKAGHL 187
P.chrysogenum 124 -HS--SN-ED-GRSI---PIPLDEK----K-ARQ-QHSHESDSGLGTSVSSEEGLADSK--NK 170
U.reesei 140 -ID--TA-PYDSKFH---SSR-RLP----P-VDA-KEHRVADSGIGSSILDAPLEKTP----- 183
A.capsulatus 147 DVD--SI-QKDKAFK---PV----R----D-QDR-RRRLSCDSGIGSSISGASMSSSVDTREQ 193
C.immitis 141 -VD--TA-TNHARPN---RSR-RLP----P-VDI-KHRRTSDSGIGSSIGEALPHKSAAGSGQ 189
A.benhamiae 141 -ID--SS-FR-NQVQ---TG---PP----G-KEN-LRAELSDSGVGSSVGEDILKRE------ 180
A.gypseum 141 -ID--SS-FR-NSTE---SG---FP----G-KEN-IRTELSDSGVGSSIGEDVLNRERAGAVE 186
T.rubrum 141 -ID--SS-FR-NQVQ---TG---PP----G-KEN-LRAELSDSGVGSSVGEDILKRE------ 180
P.brasiliensis 162 -TD--ST-DS-NHCK---PLRIDDS----P-AMN-SSQHASDSGLGTSVSGTSNSASIHSREK 210
P.nodorum 129 -----LD----------S--KPSA----V-QRN-THEHASDSGLGSSIGGSKYGKHAARTQ- 166
P.tritici-repen 130 -----LD----------V--DVFK----S-AAA-EHDHASDSGLGTSIGESKYDDIRTQ-R- 166
C.globosum 122 -QP--AE-R--R------M--LRPR----T-RRS-SRFHESDSGLGTSIASTNDKDAVAKKD- 163
N.crassa 114 -VE--PE-R--R------M--LRPR----T-RRS-SRHHDSDSGLGSSIASTSEKDAS-SKA- 154
P.anserina 122 -KE--VE-R--K------V--LRPR----T-RRT-SRHHDSDSGLGTSIASTNEKIAA-KEQ- 162
S.macrospora 114 -AE--NE-R--R------M--LRPR----T-RRS-SRHHDSDSGLGSSIASTSEKDAS-CKA- 154
M.grisea 117 -----VE-R--R------V--LRPR-PVRR-PGQ-SHHHESDSGLGSSILSTKQKSTP-NAD- 158
T.reesei 126 ----EPE-R--R------V--LRPR----S-VRR-TRNHASDSGIGSSVVSTNDKAGAADST- 166
G.zeae 127 -----EE-R--R------T--LRPR----GRTSR-ALEDHSDSGLGTSVASTNEKRGAVTAS- 167
N.haematococca 131 -----AE-R--R------V--LRPR----S-VRR-SRHHASDSGLGTSVASTNEKHAAVTQS- 170
V.albo-atrum 124 -TKTSGE-R--R------V--LRPR----H-RRS-SEQHASDSGLGTSLASSVEKQAPSITS- 167
B.cinerea 133 GVM-PSEQS--S------V--RRST----R-PRR-TTPRYADSGLGSSIGSSSSDKKVLTND- 177
cons 190 .***:*:*: 252
A.nidulans 177 -GQ--------L-SRSNLPT--SRSQSAITRSISAM-DAQSTQRHKLSSEGRAEIEKHVIGPL 226
A.fumigatus 185 AGQ-LSFES---QSSAKATT--RMTQSAITSSISAFD-AKISQKRQLGLPACKQIERLIIKPI 240
N.fischeri 185 AGQ-LSFES---QSSAKAAT--RMTQSAITSSISTFD-AKVSQKRQLGLPACKQIERLVIKPI 240
A.clavatus 180 AGQ-LSFEP---QVSGMSGS--QMTQSAITSSISAIN-PKQIAKYQLGARACKQIEKNLLVPI 235
A.oryzae 179 ----------NVTAGQVCQ----ASVGHKLSATARVEIQGRILFPL 210
A.flavus 179 ----------NVTAGQVCQ----ASVGHKLSATARVEIQGRILFPL 210
A.niger 172 ----------SR-SKTV--TRTQSAITSSLMAV-EPDSASKQNLSLGASQEIEKRILTPL 217
A.terreus 175 ----------ESADLDSNAPM-S--ASHLRKISPDGVDKIEHRVFYPL 209
T.emersonii 188 -C----------TGNADL--PGGQSAITHSISSGD----TSRRQLPLAACKQIERFILAPI 231
P.chrysogenum 171 V----------NDGI--HDNQSAITSSISAFD-TGSSPKRQLGPAACKQIERFVLVPI 215
U.reesei 184 ----------ASGGPLMRFGAFKQIERHILAPI 206
A.capsulatus 194 VRK-ATHSF-----QEEGKI--VSTQSAITRSISSTT-AL-AQIPSFSVTARKHIDRFILVPI 246
C.immitis 190 ----------PRK--SPSRSAITRSISYKSSTMDASKPVLSPTAVKQIERRILIPI 233
A.benhamiae 181 ----------AR--ESVV----SSSSTLSSGGKRHIQRHILLPL 208
A.gypseum 187 ----------VSTR--ESVA----SSRTKLSTGAKRHIQRHVLLPL 216
T.rubrum 181 ----------AR--ESVV----SSSSTLSSGGKRHIQRHILLPL 208
P.brasiliensis 211 AIDKMNHSPKDRKSLQEPKM--ASTQSAITRSISSNT----LEIPTFGVAARKNIDRFILNPI 267
P.nodorum 167 PSS-------VS-DSVNTSI--TSTHSAITRSMSSL-GA-SEEKHTLGEYACQQIHDNIIKPI 217
P.tritici-repen 167 -SS--------R-ESISSSI--SSTHSAITRSFAAL-GA-SDEDHTLGEYACKQIHDSIIKPI 215
C.globosum 164 ----------K---AGRTTAVTRSAAAR-SATAASAVGLSPRATSRIYAHTLKPL 204
N.crassa 155 ----------K---TTRTSAVARSATAR-AASTPDLPGLGDRATNRIVEYILKPL 195
P.anserina 163 ----------T---VAKTTAVTRSAAATRTTTTTTTQVLGQRANNRICEHTLKPL 204
S.macrospora 155 ----------K---TTRTSAVTRSAAAR-SASALNLPGLSDRATNRIVEYVLKPL 195
M.grisea 159 ----------STPAK--TGVKGSAVTRSAAS---STSKNLPSLSARATNRVFEHTLKPL 202
T.reesei 167 ----------KK---PQASALTRSAASS-T--TAMLPSLSHRAVNRIREHTLRPL 205
G.zeae 168 ----------KE-AKVQTRCLTRSAAAA-AA-TGKLPSLGSKAFSRIHEHTLRPL 209
N.haematococca 171 ----------KE-T-KTRSAITRSAAAS-S--AEKLPSLGSKAINRIHEHTLRPL 210
V.albo-atrum 168 ----------KT---SKASAITRSAAAP-SNTMTKVSGLSSKAVSRVHEHVLRPL 208
B.cinerea 178 ----------NVSKSTKPSTITRSAAAP-STTIKSLPRLSARTTARIQEHILKPL 221
cons 253 : .: : *: 315

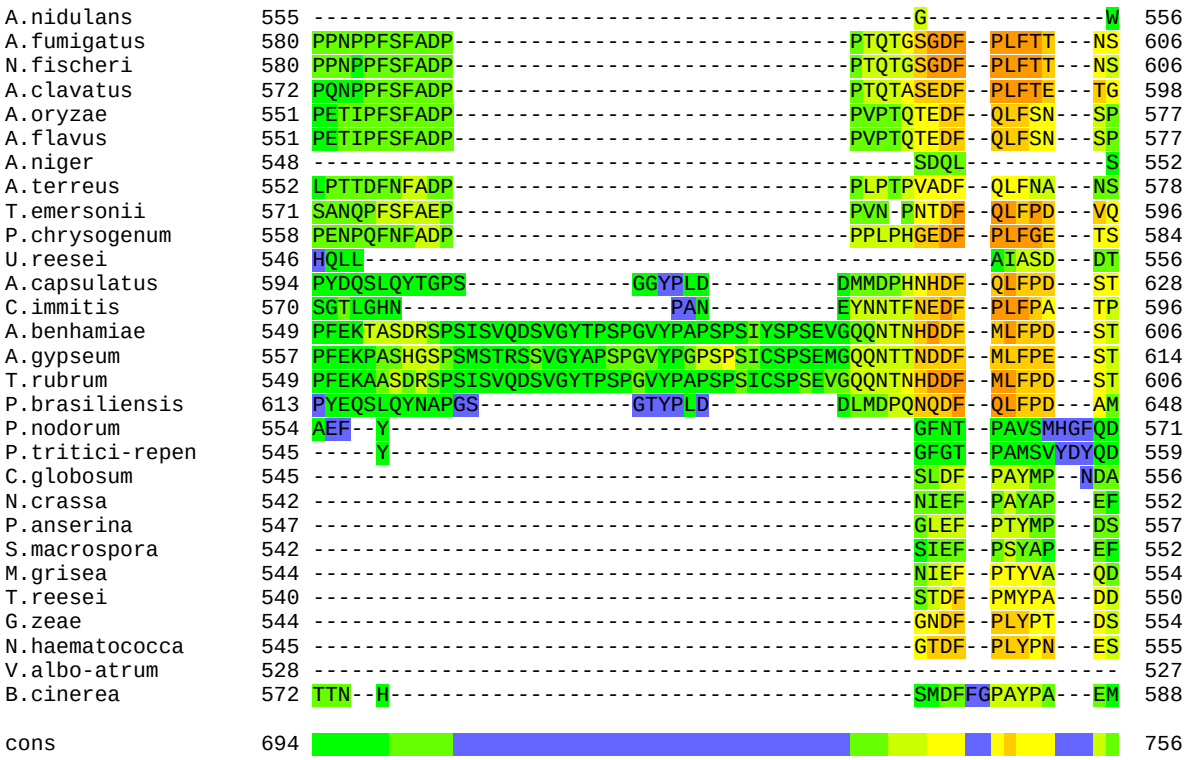

Intron 2
A.nidulans 227 LEDEKSKPFHPILEDVRQQIDDERISCLRDLEKTVFSLAP-EVKTNDAAYVRFCQYTILCLGQ 288
A.fumigatus 241 LREERLKPFHPLVQSIPQRIVDREIGCLRDLEKTLLWLAP-NYAASRASYLRFCEFTIQCLHT 302
N.fischeri 241 LKEERLKPFHPLVQSIPQRIVDREIGCLRDLEKTLLWLAP-NYAASRASYLRFCEFTIQCLHT 302
A.clavatus 236 LGEEKLKAFHPLVRSVPQRIVSKEINCLRDLEKTLLWLAP-HYATTRASYIGFCEFTIQCLYS 297
A.oryzae 211 LMKDKFQLFHPLVRCAHQQIEKQQLKCLRDVEKTLLFSTP-DVKAPTAAWYSFCRFTIHCLHE 272
A.flavus 211 LMKDKFQLFHPLVRCAHQQIEKQQLKCLRDVEKTLLFSTP-DVKAPTAAWYSFCRFTIHCLHE 272
A.niger 218 LGQELCKPFHHLVESAREQIEKKQIGTLRDLEKSLLHGAA-DVEAEVGSYYQFCSSTILCLSE 279
A.terreus 210 LMNTQFDKFHSTVRYAAQGVKDNRFRCLRDVENLFRNPFPIDHVNEKEEFSLFLNQVVLCLDD 272
T.emersonii 232 LREERLKPFHPLVRSIPQRVVDKEITCLRDLEKTLLWLAP-QYSITKASYLGFCEFTIQSIHT 293
P.chrysogenum 216 LKEPRLKPFHPLVRSVPQRIINKQIVCLRDLEKTLLWLAP-NSATSRNSYLNFCEFTIQCLHT 277
U.reesei 207 LGEQSFKPFHPIVETIPQRVRNNNITCLRDLEKILLFVPL-NETISKGSYMQYGRFTIHCLHT 268
A.capsulatus 247 LREKRLSPFHPLIRSVPQRIESKEIACLRDLEKTILFLAP-RYTQSKAVYLGFCEFTIQCLHT 308
C.immitis 234 LREKHLGHFHPLVRGVPQRIKLNDITCLRDLEKTLLFLAP-DFAPTRTSYISFCDFTIQCLHT 295
A.benhamiae 209 LKEQRLKNFHSLVRSVPQRIQSSEITCLRDVEKTLLYLAP-HHTISRSSYLSFCEFTIQCVHT 270
A.gypseum 217 LKEQRLKTFHSLVRSVPQRIQSSEITCLRDVEKTLLYLAP-HHTVSRSSYLSFCEFTIQCVHT 278
T.rubrum 209 LKEQRLKNFHSLVRSVPQRIQSSEITCLRDVEKTLLYLAP-HHTISRSSYLSFCEFTIQCVHT 270
P.brasiliensis 268 LLEKRLTPFHPLARGVPERIENKEITCLRDLEKTLLFLAP-KFTTSKASYIAFCEFTIQCLHT 329
P.nodorum 218 LAEESLKDFHPLIQDVPRRIGEKSICNLRDLEKTLIFLAP-ELSASATSYLQFCELSIQLLHA 279
P.tritici-repen 216 LAEDSLKDFHSLVKDVPRRIGEKNIRNLRDLEKTLIFLAP-ELSATSDSYSRFCQRSIQLLAT 277
C.globosum 205 LANSGLKDFHPLLLECPRKIQERQIVCLRDLEKTLLLVAP-ERTKSSELYLDFCLTTIRCIQA 266
N.crassa 196 LAKPNLKEFHSLVLECPKKIQEKEILCLRDLEKTLILVAP-ERTKAAGSYLDFCLTTIDCLQA 257
P.anserina 205 LGKPEFKDFHPLLLECPKKIQDKEIVCLRDLEKTLLLVAP-ERTKSAGLYLDFCLTTIQCIQA 266
S.macrospora 196 LAKPALKEFHSLVLECPKKIQDKEILCLRDLEKTLILVAP-ERTKAAGSYLDFCLTTIDCLQA 257
M.grisea 203 LADSSFKEFHPVLLECPRKIQSKQIVCLRDVEKTLLLTAQ-GVTKVPKLYLDFCLSTVQCIRA 264
T.reesei 206 LEKPTLKEFEPIVLDVPRRIRSKEIICLRDLEKTLIFMAP-EKAKSAALYLDFCLTSVRCIQA 267
G.zeae 210 LAKPTLKEFKPIVLDIPRRIQSKEIICLRDLEKTLIFMAP-EKAKSATLYLDFCLTSVRCIQA 271
N.haematococca 211 LAKPTLKDFEPIVLDIPRRIRSKEIICLRDLEKTLIFMAP-ERTKSAALYLDFCLTSVRCIQA 272
V.albo-atrum 209 RAKPELKDFEPIVLDIPRRIRDKEIICLRDLEKTLIFMAP-ERAKTAALYLDFCLTSIRCIQA 270
B.cinerea 222 LAKESFKDLHPLVQDCPRRIHQKEIVSLRDLEKTLIFTAP-KYTKSASLYLDFCLESIHCIQA 283
cons 316 . :. . : : ***:*: . : : : : 378

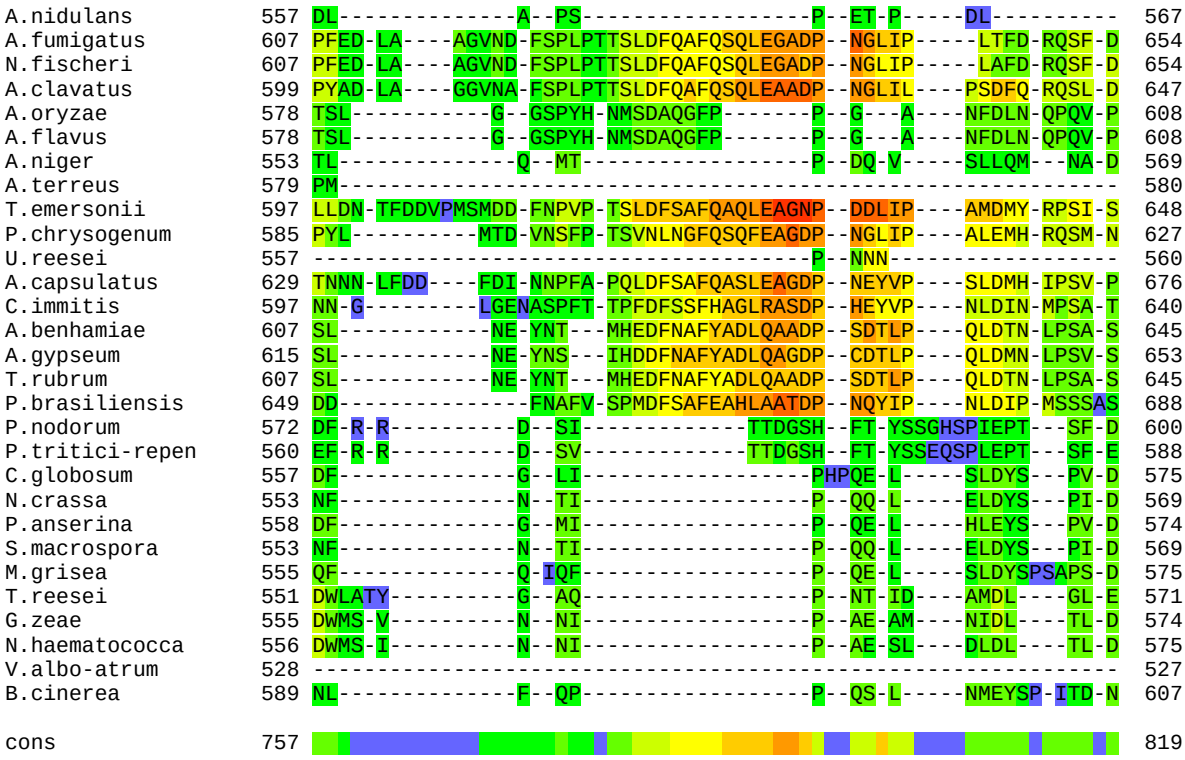

Intron 3
A.nidulans 289 TVSFLNGRDLCLPTDKQYNNGYFVDLLDQVSQFKRIRDEWKRRHE--------A-D---GKVK 339
A.fumigatus 303 SVYHLNERDQRLPADRPYTNGYFLDLVAQIRRYAAMINESRSSMPSN---REPAQNGAKASAP 362
N.fischeri 303 SVYHLNERDQRLPADRPYTNGYFLDLVAQIRRYAAMINESRSSMPSN---REPAQDGAKASAP 362
A.clavatus 298 TVQHLNSRDQRLPTDRPYTNGYFLDLVAQIHQYAAMINASRNQMSSN---E-KP-AR-ASASS 354
A.oryzae 273 TSGYLRGRDLTLPNDVPYSNNYFLDLITQINRFARIRDATRSRQES----ATNG-EK-AAKSL 329
A.flavus 273 TSGYLRGRDLTLPNDVPYSNNYFLDLITQINRFARIRDATRSRQES----ATNG-EK-AAKSL 329
A.niger 280 TYTHLDPRDLCIPTDKSYSNSYFLDLTAQVHRFKAMRDEARKNKK---------------EAS 327
A.terreus 273 TWHRLDESDRTMPGDVPYSNEYILDLYDQIVRFKALCERAKQNLS------QKD-SN-GATSK 327
T.emersonii 294 AVTHLNERDQRLPTDRPYTNGYFLDLVAQIRQYAAMIAASRERARAR---GEQA-DA-ARHNS 351
P.chrysogenum 278 SASHLNDRDQRLPADRPYTNGYFLDLVSQVRRYAAMVRAERERVQAT---QGSESNKETKSPI 337
U.reesei 269 AVRHLNDRDQCLPADRPYTNGYFLDLVAQVQGYAATIAAARNSRSSR---DQKG-NN-LDYSS 326
A.capsulatus 309 TVGYLNDRDQRRPSDRPYTNGYFLDLVEQMRQYAALVGATRERQGPRPRSRIAE-NN-MEYSS 369
C.immitis 296 TVGFLNDRDQRRPADRPYTNGYFLDLVEQVRQHAVMLHSTRQSQG------KNE-NN-LDYSS 350
A.benhamiae 271 TVGYLNDHDQRRPADRPYTNRYFLDLVEQIRQYAALIRSSRERRAT----SSNR-DE-LDYSS 327
A.gypseum 279 TVGYLNDHDQRRPADRPYTNGYFLDLVEQVRQYAVLIRTARERRAA----SPNR-DD-LGYSS 335
T.rubrum 271 TVGYLNDHDQRRPADRPYTNGYFLDLVEQIRQYAALIRSSRERRAT----ASNS-DE-LDYSS 327
P.brasiliensis 330 TVGYLNERDQRRPSDRPYTNGYFLDLVEQIRQYASMISAARERQGPRPRSRDGE-NN-KDYSS 390
P.nodorum 280 TVEQLSEQDQRLPSDRPYTNNYFLDLAEQIQRYARIMAATRQKESK----GESL-DD-MDYSR 336
P.tritici-repen 278 TVQYLSEQDQRLPSDRPYTDGYFLDLMEQIRRYASIMAATREKEVK----GEDL-DD-MDYSR 334
C.globosum 267 TVEYLSDREQTRPRDVPYTSGYFVDLVDQIRNYAQQLADAKENQ----------DE-MDVDP 317
N.crassa 258 TVQHLGDRELTRPRDLPYTSGYFVDLVDQFYNYARQIAESNKTKE-------GA-ND-MDIDP 311
P.anserina 267 TVEYLSDREQTRPRDVPYSSGYFIDLVDQIRHYAQQLSEAKEKG--------EN-DE-MDVDP 319
S.macrospora 258 TVQHLGDRELTRPRDLPYTNGYFVDLVDQFYNYARQIAESNKTKD-------GA-NG-MEVDP 311
M.grisea 265 TVVYLSETEQRRPKDVPYSSGYFIDLVDQIKQYAQQLADQKKAG----------A-SDPNI 314
T.reesei 268 TVEYLTDREQVRPGDRPYTNGYFIDLKEQIYQYGKQLAAIKEKGS-------LA-DD-MDIDP 321
G.zeae 272 TVEYLSDREQIRPADRPYTNGYFLDLKDQILEYGKQLAAKNS----------G-DE-MDIDA 321
N.haematococca 273 TVEYLSDREQIRPADRPYTNGYFIDLKDQILEYGKQLAAKNN----------G-DD-MDVDA 322
V.albo-atrum 271 TVEYLSDREQIRPADRPYTNGYFIDLVEQIRQYAGQLATAKEAGV-------EG-RE-MDVDP 324
B.cinerea 284 TVELVNEREQTRPNDRPYTNGYFVDLVEQIKQYAEQVREAKEKEEK----GETI-SE-MDAHP 340
cons 379 : : : * * *.. *::** *. . . 441

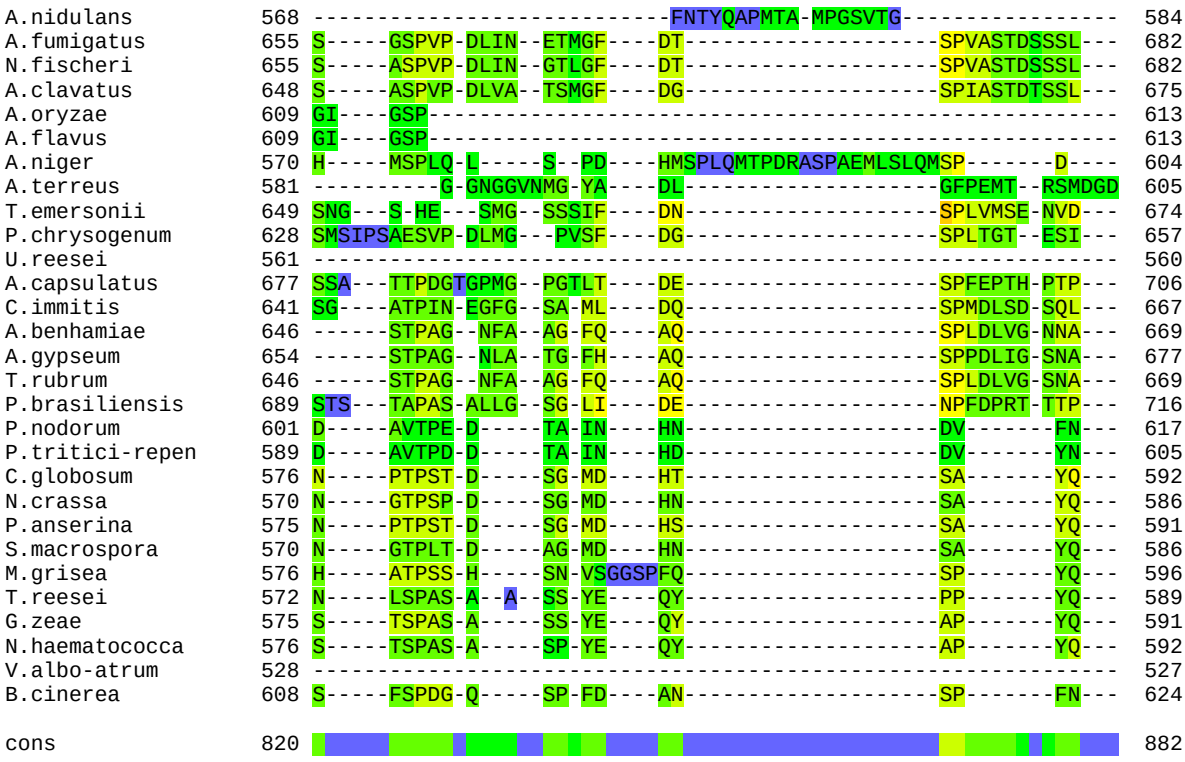

Region
A.nidulans 340 APQLRLEGGLSQTGRLLEMVV--EQDGEAISLRTGKPYEG-----QPIPSMKRSLSAAS-TD- 393
A.fumigatus 363 NEHITLEGGLSKNGRPAELVVH--KDGEMISLRTGKPYEENA-----VPAMKRSLSLGS-VD- 416
N.fischeri 363 LEHVTLEGGLSKNGRPAELVVH--KDGEMISLRTGKPYEEDA-----VPAMKRSLSLRS-VD- 416
A.clavatus 355 SDPVTLEGGLSKNGRPAELVVH--KDGEMISLRTGKPYEEGA-----TPSIKRSLSIGS-VD- 408
A.oryzae 330 EPRLTLEGGMSETGRPAELVMH--KDGKSISLQTGKPYDEH-----AIPTFKRTLSVET-VD- 383
A.flavus 330 EPRLTLEGGMSETGRPAELVMH--KDGKSISLQTGKPYDEH-----AIPTFKRTLSVET-VD- 383
A.niger 328 ELKLTLEGGMSQTGRPLELVA--QKEGQAVSMQTGEAYDA-----HAPPLVKRTLSLNE-AD- 381
A.terreus 328 LPKLVFKGGLAKTGRQAELVAQ--SDDQMMSLRTGKPYQES-----QVPSMKRGLSLCS-TDG 382
T.emersonii 352 EERLTLEGGLSQTGRPAELVCH--KDGKAISLRTGEPYDEKA-----VPTMKRTLSNQS-ID- 405
P.chrysogenum 338 IPKATLEGGLSQNGKSAELVVM--QDGKPISMATGKPYEADI-----PGVAKRTISLDDAVD- 392
U.reesei 327 GEELVLEGGLSSSGRPAELVRR--KKDVAISLQTGEPYVESK---PTIPIMKRALSVEA-GD- 382
A.capsulatus 370 DEEVTLEGGLGATGRPAELVRR-KKDKQAISLRTGNPYDEKA--SPAVPIMKRSLSMES-CD- 427
C.immitis 351 GEELHLEGGLSSTGRPAELVRR--KNGVAISLKTGEPYVESK---PIMPAMKRSLSIDS-GD- 406
A.benhamiae 328 GEELVLEGGLASTGRPAELVRK--KKGQAISLKTGEPFVEQK---PEVPSLKRAPSMEN--Q- 382
A.gypseum 336 GEELVLEGGLASTGRPAELVRK--KKGQAISLKTGEPFVEQK---PVVPSLKRAPSMEN--Q- 390
T.rubrum 328 GEELVLEGGFARTGRPAELVRK--KKGQAISLKTGEPFVEQK---PEVPSLKRAPSMEN--R- 382
P.brasiliensis 391 DEELTLEGGLAATGRPAELVRR--KKNQAISLRTREPYEERA--SPVIPMMKRSLSTES-CD- 447
P.nodorum 337 DEKITLRGGFSHDGRPMQLVR--EKNGKSMSLAEGIDS--------SYKVSKRPLSDDE-MDE 388
P.tritici-repen 335 DEKITLRGGLSHNGRPVELVR--EKNGKVVPLADS-------------PTIKRRFSDED-EDE 381
C.globosum 318 TDEIKLHGGIHINGRPAELVRV-KKNGKMISMATGQPIDKIDEEADGTVRFKRSASQEL-EDE 378
N.crassa 312 TDQIKIHGGPHINGRLSELVRV-KKNGQAISLATGLPVDLCDKAPETPVNFKRSQSEEA-LDE 372
P.anserina 320 TDEIKLHGGIHINGRPAELVRI-KKNGKMISMATGEPIESIEEESSGAVRIKRSASEEL-EDE 380
S.macrospora 312 TDQIKIHGGPHINGRLSELVRV-KKNGQAISLATGLPVDLCEKAPESPVNFKRSQSEEA-LDE 372
M.grisea 315 KEEVKLHGGIAVNGRPAELVR--ISGGKAISLATGEPVELDE-ETTSPIRFKRSASQQL-ADD 373
T.reesei 322 SDEVRLYGGVAENGRPAELIRV-KKDGTAYSMATGKIVDMT----ESPTPLKRSLSEQR-EDE 378
G.zeae 322 SDEIKLVGGLSVNGRPAELVRV-RKDGTYISLDTGKPVETDD---DAPMKMKRSLSQQL-EDE 379
N.haematococca 323 SDEIKLIGGIAENGRPAELVRV-RKDGTYISLETGEPVEVD----EAPMKMKRSLSQQL-EDE 379
V.albo-atrum 325 TDEVKLFGGISQNGRPAELVRV-RKDGQAISMATGLPVDMDE-DGKDFPRLKRSLSQQL-ADD 384
B.cinerea 341 SDEVKLHGGLTRNGRPAELVRINKKTGKAISIATGQPVDMDD-DDNKSMRFKRSLSEEA-EDE 401
cons 442 504

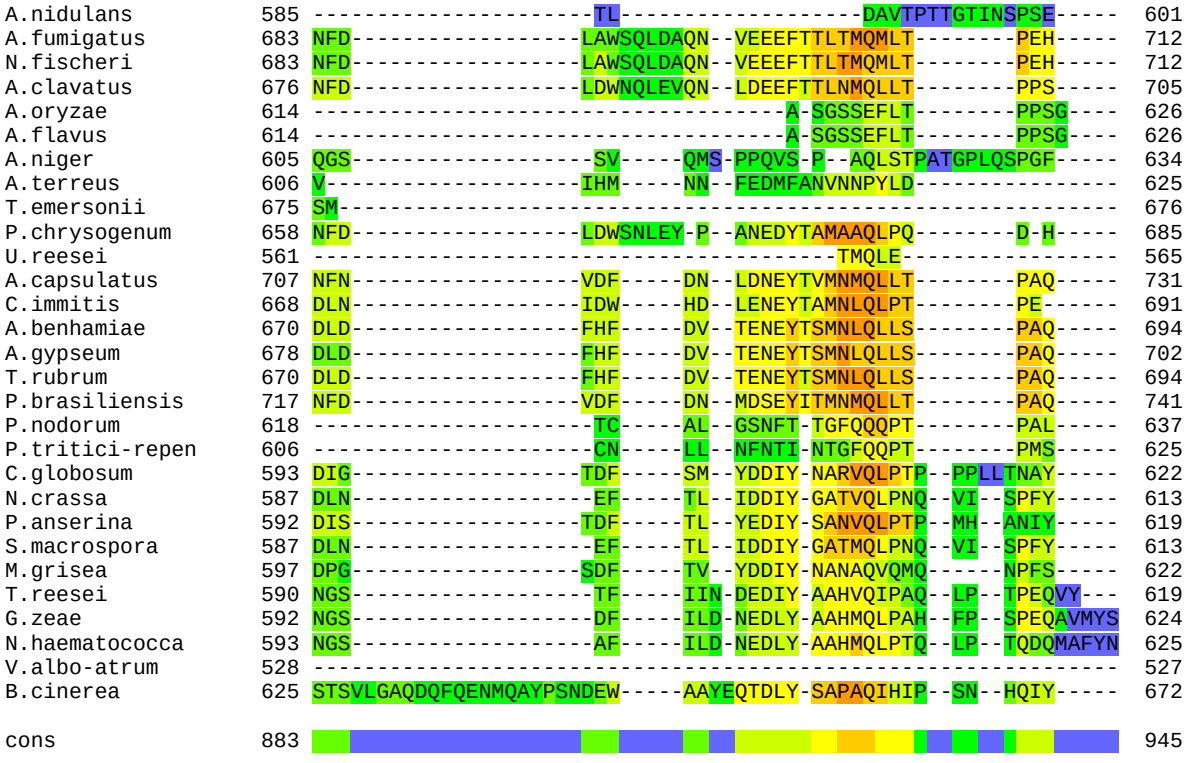

Zinc finger 1
Zinc
of putative NLS
Intron 4
A.nidulans 394 EGVQRSMARRKKNAPPMNIN-KKCK--DCDKVFARPCDLTKHEKSHSRPFKCPVTSCKYHIKG 453
A.fumigatus 417 EGVERSMARRKKNAPPMDIN-QKCK--DCDKVFKRPCDLTKHEKTHSRPWKCTEPSCKYHEIG 476
N.fischeri 417 EGVERSMARRKKNAPPMDIN-QKCK--DCDKVFKRPCDLTKHEKTHSRPWKCTEPSCKYHEIG 476
A.clavatus 409 EGVERSMARRKKNAPPMDIN-QKCK--DCDKVFKRPCDLTKHEKTHSRPWKCNDDSCKYFEVG 468
A.oryzae 384 EGVERSMARRKKNAPPMDIN-QKCQ--FCDKVLKRPCDLTKHEKTHSRPYKCPERGCKYFELG 443
A.flavus 384 EGVERSMARRKKNAPPMDIN-QKCQ--FCDKVLKRPCDLTKHEKTHSRPYKCPERGCKYFELG 443
A.niger 382 EGARRSMARRKKNAPPMNIN-TKCS--HCDKIFQRPCDLTKHEKTHSRPFKCPFEGCKYHELG 441
A.terreus 383 EGAHRCMARRRKNEAPMNIN-TPCD--FCGQIFARPCDLTKHMKTHTRPFKCSVPECKYYTYG 442
T.emersonii 406 EGVARSMARRKKNAPPMDIN-QKCR--DCDKVFKRPCDLTKHEKTHSRPWKCTEPTCKYFEIG 465
P.chrysogenum 393 EGVERSMARRKKDAPPMNIN-QKCA--DCDKVFKRPCDLTKHEKTHTRPWKCEYTDCSYHTKG 452
U.reesei 383 DSVMRSMARRKKNEPPLNIN-KKCD--HCDRIFRRPCDLSKHEKTHTRPWKCTEPGCKYSKTG 442
A.capsulatus 428 DGVARSMARRKKNAPPLDIN-QKCK--DCDKVFKRPCDLTKHEKTHSRPWKCAEKSCKYFEIG 487
C.immitis 407 DGVMRSMARRKKNEPPLNIN-KKCD--HCEKVFRRPCDLTKHEKTHTRPWKCTEPTCKYSKVG 466
A.benhamiae 383 ESVMRSMARRKKDAPPMNIN-EKCK--HCDKVFRRPCDLTKHEKVHTRPWKCLDTKCKYYQLG 442
A.gypseum 391 ESVMRSMARRKKDAPPMNIN-EKCK--HCDKVFRRPCDLTKHEKVHTRPWKCLDTKCKYHQLG 450
T.rubrum 383 ESVMRSMARRKKDAPPMNIN-EKCK--HCDKVFRRPCDLTKHEKVHTRPWKCLDTKCKYYQLG 442
P.brasiliensis 448 DSVTRSMARRKKNAPPLDIN-QKCK--DCDKVFKRPCDLTKHEKTHSRPWKCDERSCKYFEIG 507
P.nodorum 389 DEATRSMARRRKSDKPGDVM-HTCR--DCKKEFKRPCDLTKHEKTHSRPWKCSEEKCKYFDLG 448
P.tritici-repen 382 DEVTRSMARRRKSEKPGDVM-HTCR--DCKKQFKRPCDLTKHEKTHSRPWKCSEKNCKYYDLG 441
C.globosum 379 EEIMRSMARRKKNASPEELAPKKCREPGCNKEFKRPCDLTKHEKTHSRPWKCPVPTCKYHEYG 441
N.crassa 373 EEVMRSMARRKKNASPEELAPKRCREPGCNKEFKRPCDLTKHEKTHSRPWKCPVKTCKYHEYG 435
P.anserina 381 EEIMRSMARRKKNATPEELAPKKCREHGCNKEFKRPCDLTKHEKTHSRPWKCPVTTCKYHEYG 443
S.macrospora 373 EEVMRSMARRKKNASPEELAPKRCREPGCNKEFKRPCDLTKHEKTHSRPWKCPVKTCKYHEYG 435
M.grisea 374 DEIMRSMARRKKNAPPEEYAPKMCREPGCGKEFKRPCDLTKHEKTHSRPWKCPVPTCKYHEYG 436
T.reesei 379 EEIMRSMARRKKNATPEDVAPKKCREPGCTKEFKRPCDLTKHEKTHSRPWKCPIPTCKYHEYG 441
G.zeae 380 EEIQRSMARRKKNASPEELAPKKCREPGCNKEFKRPCDLTKHEKTHSRPWKCPVSTCKYHTYG 442
N.haematococca 380 EEIQRSMARRKKNASPEELAPKRCREPGCNKEFKRPCDLTKHEKTHSRPWKCPVPTCKYHTYG 442
V.albo-atrum 385 EEIMRSMARRKKNATPEELAPKKCREPGCNKEFKRPCDLTKHEKTHSRPWKCPVKTCKYHEYG 447
B.cinerea 402 ESVMRSMARRKKDASAAELAPKFCREKGCDKSFKRPCDLTKHEKTHSRPWKCPVTSCKYHEYG 464
cons 505 567

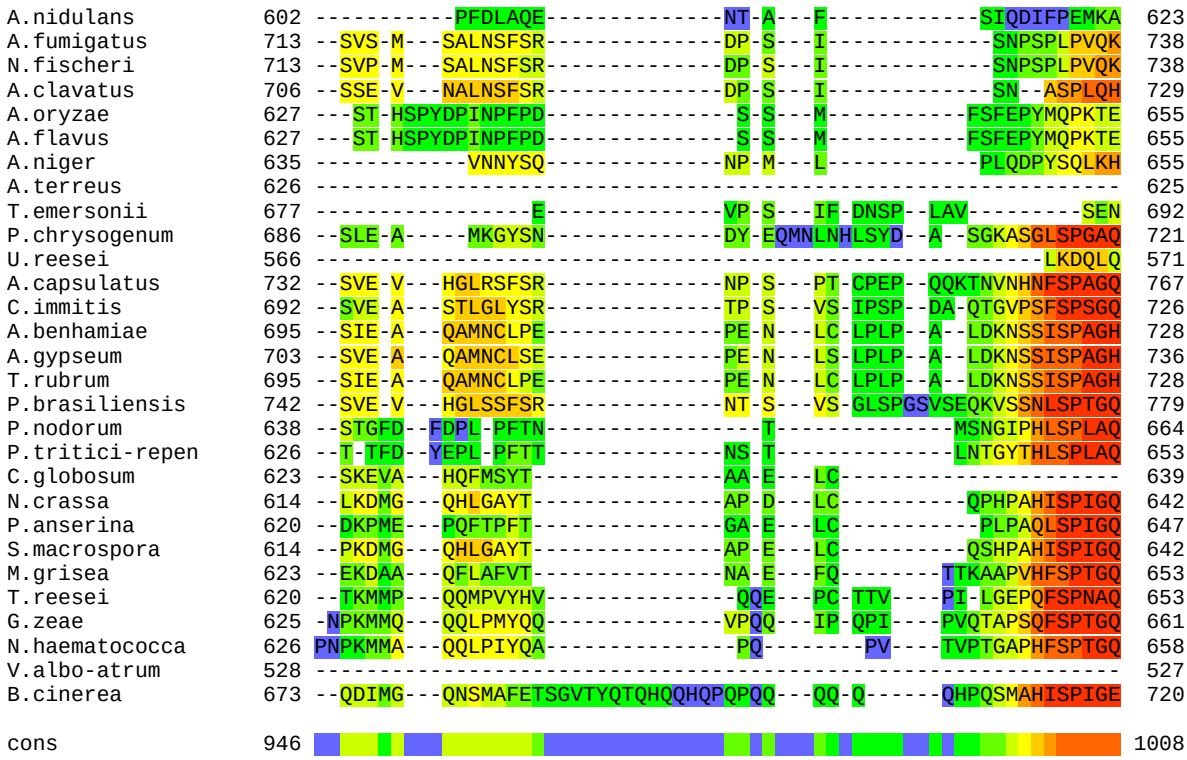

finger 2
Zinc finger 3
PTC STOP in A.nidulans sltA1 strains
A.nidulans 454 WATEKESERHYNDKHSDAPRLFACQFESCSYKSKRESNCKQHMEKTHGWVYMRSKNNGRSKAS 516
A.fumigatus 477 WPTEKERDRHINDKHSKAPALYKCKFAPCTYSSKRESNCKQHMEKAHGWDYVRSKHNGRNSKK 539
N.fischeri 477 WPTEKERDRHINDKHSRAPALYKCKFAPCTYSSKRESNCKQHMEKAHGWDYVRSKHNGRSSKK 539
A.clavatus 469 WPTEKERDRHINDKHSKAPALYKCTFAPCSYQSKRESNCKQHMEKAHGWVYVRSKHNGRNSKK 531
A.oryzae 444 FPTEKETERHYNDKHCKNPRLFRCHQPGCTYASKRESNCKQHMEKTHGWVYERTKNNGKNKPV 506
A.flavus 444 FPTEKETERHYNDKHCKNPRLFRCHQPGCTYASKRESNCKQHMEKTHGWVYERTKNNGKNKPV 506
A.niger 442 WPTEKENERHVNDRHSTTPRMYACTFNGCAYKSKRESNCKQHMEKAHGWNYVRAKNNGRNAKR 504
A.terreus 443 FPTEKEKDRHFNDKHNPDPEPYECDLGGCNYRSKRLSNLKQHKEKKHGWQYVRTKSNGKRKDK 505
T.emersonii 466 WPTEKERDRHMNDRHSTSPALFKCKFAPCTYQSKRESNCKQHMEKAHGWVYVRSKNNGKNRKS 528
P.chrysogenum 453 WPTEKERDRHMNDRHSDKPTLYKCQFHKCPYASKRASNCKQHMEKSHGWVYVRSKNNGRNGSK 515
U.reesei 443 WPTEKERDRHVNDKHCKSPRLFNCRFPPCTYQSKRESNCKQHMEKAHGWVYVRSKNTGKGSDR 505
A.capsulatus 488 WPTEKERDRHVNDKHSKSPPLFKCHFSPCTYQSKRQSNCKQHMEKAHGWVYIRSKNNGKSGSR 550
C.immitis 467 WPTEKERDRHVNDKHSKSPPLYSCLFKPCTYQSKRESNCKQHMEKAHGWVYVRSKNTGKASAR 529
A.benhamiae 443 WPTEKERDRHMNDRHSKSPKKYKCLYPGCTYDTKRQSNCKQHMEKLHGYKYVRSKNNGRASHQ 505
A.gypseum 451 WPTEKERDRHMNDRHSKSPKKFKCLYPGCTYDTKRQSNCKQHMEKLHGYKYVRSKNNGRSSHQ 513
T.rubrum 443 WPTEKERDRHMNDRHSKSPKKYKCLYPGCTYDTKRQSNCKQHMEKLHGYKYVRSKNNGRASHQ 505
P.brasiliensis 508 WPTEKERDRHVNDKHSKSPQLYRCQFPPCTYQSKRDSNRKQHMEKAHGWVYVRSKNNGKASSR 570
P.nodorum 449 WPTEKERDRHMNDKHSAQPAQYKCLYPPCTYASKRESNCKQHMEKAHGWEYVRSKSNGRKKMP 511
P.tritici-repen 442 WPTEKERDRHMNDKHSAAPAQYKCLYPPCTYASKRESNCKQHMEKAHGWEYVRSKCNGRKKGP 504
C.globosum 442 WPTEKEMDRHNNDKHSASPPMYECLFKPCPYKSKRESNCKQHMEKAHGWVYVRTKTNGKKDSS 504
N.crassa 436 WPTEKEMDRHHNDKHSSAPPMYECLFKPCPYKSKRESNCKQHMEKAHGWTYVRTKTNGKKPST 498
P.anserina 444 WPTEKEMDRHHNDKHSAAPPMFECYYKPCPYKSKRESNCKQHMEKAHGWTYVRTKTNGKKPGS 506
S.macrospora 436 WPTEKEMDRHHNDKHSSAPPMYECLFKPCPYKSKRESNCKQHMEKAHGWTYVRTKTNGKKPST 498
M.grisea 437 WPTEKEMDRHHNDKHSAAPPMYECLYKPCPYKSKRESNCKQHMEKAHGWTYVRTKANGGKKID 499
T.reesei 442 WPTEKEMDRHINDKHSDAPAMYECLFKPCPYKSKRESNCKQHMEKAHGWTYVRTKTNGKKAPS 504
G.zeae 443 WPTEKEMDRHHNDKHSAAPAMYECSFKPCPYKSKRESNCKQHMEKAHGWTYVRTKTNGKKLPS 505
N.haematococca 443 WPTEKEMDRHHNDKHSDAPAMYECLFKPCPYKSKRESNCKQHMEKAHGWHYVRTKTNGKKLPG 505
V.albo-atrum 448 WPTEKEMDRHHNDKHSSAPPMHECLFQALPLQVEARVNCKQHMEKAHGWQYVRTKTNAAEAPA 510
B.cinerea 465 WPTEKEMDRHQNDKHSAAPPLFECHFKPCPYRSKRESNCKQHMEKAHGWEYIRSKNNGKNRPA 527
cons 568 :.**** :** **:* * . * : * *** ** **: * *:* .. 630
A.nidulans 517 ---PQ--Q--QTTSPSSSSVQPKQAP----S---V-WSMTPPSEAPDYR---QEPN------- 554
A.fumigatus 540 ---A-----SNGATPQTPS---IATPS-SKA---Q-GITTPLTGS--EPSPF--EPVTAY--- 579
N.fischeri 540 ---A-----SNGATPQTPS---IATPS-SKA---Q-DIATPLTGS--EPSPF--EPVTAY--- 579
A.clavatus 532 ---A-----STRAASQTPS---IATPS-SKA---Q-EILTPATRP--TPSPF--ESVASY--- 571
A.oryzae 507 ---KQ-GS--AQPTPQSSG---IPSPA-ASHP-TG-DFSTPTTGP--TVSPS--EPPIAF--- 550
A.flavus 507 ---KQ-GS--AQPTPQSSG---IPSPA-ASHP-TG-DFSTPTTGP--TVSPS--EPPIAF--- 550
A.niger 505 ---RAT-S--ARASPAEESTA-HSSPL-HMS---P-DQVSPVQLAP-AQLASLQMA------- 547
A.terreus 506 ---VKGKKASPQSTPDTPG---LTTPG-TSTA--Q-SFSTPNTGP--SPSPP--QAVSRP--- 551
T.emersonii 529 ---ASVS--SAQPTPQTPA---MSTPS-NS----T-DVPTPVSGP--APSPY--DPFTGI--- 570
P.chrysogenum 516 ---R-GS--SSQATPQTPS---VSTPA-SKT---T-DFASPIPGP--SPSPS--DQNYNW--- 557
U.reesei 506 ---G-----SPH-TPPSF------NPSPTIA---P-GIPTPISTV--ACSPY--MSPYQPQLE 545
A.capsulatus 551 ---VSGST-SGQPTPRTPN---IQTPN-SGV---M-DLPTPQSHP--GRSPY--APNVS---- 593
C.immitis 530 ---G-----SNRPTPNIA------TS-PNLA---G-EMPTPISAV--APSPY--TSPYQPDFQ 569
A.benhamiae 506 ---L-----SPQPTPTTTN---ISTPQ-APS---V-DIPTPISAQ--AMSPY--LSPSEKQST 548
A.gypseum 514 ---L-----SPQPTPATTN---MNSPQ-APS---V-DIPTPISAQ--AMSPY--MSPYEQQSI 556
T.rubrum 506 ---L-----SPQPTPTTTN---ISSPQ-APS---V-DIPTPVSAQ--AMSPY--LSPSEKQST 548
P.brasiliensis 571 ---VSSS--HTPHTPQTPI---MKTPN-SGT---V-DIPTPLSHP--ALSPY--LANVS---- 612
P.nodorum 512 ---TSSER--SPTTPLTPF---LGTPM-SAV------MTTPVTPY--APSPQ--VPMMD-N-F 553
P.tritici-repen 505 ---TSSDR--SPATPLTPF---IGTPM-SAN------MTTPVTPY--IPSPQ--VPMMD---- 544
C.globosum 505 ---VNGNS--AHPTPQLQN---IPTPS-SEHS--M-GVATPPDTW--GPVYP--GS------- 544
N.crassa 499 LPSLGPDS--GHPTPQLQN---IGTPS-SDRS--M-SIATPSDDW--NAGLY--QT------- 541
P.anserina 507 S--IAGGS--THPTPQLGH---ISTPS-SDMS--A-GVATPPDDW--SH-IY--SS------- 546
S.macrospora 499 LPSLGPDS--GHPTPQLQN---IGTPS-SDRS--M-SIATPSDDW--NAGLY--QT------- 541
M.grisea 500 SN-PSGSV--THATPQLTN---MPTPP-SDNGAFA-GLATPPMEY--TMPAY--NN------- 543
T.reesei 505 ---QNGST--AQQTPPLAN---VSTPS-STPS--Y-SVPTPPQDQ-----V----M------- 539
G.zeae 506 ---IAGSV--QQQTPPLGN---MSTPS-SIEY--N-SVPTPPQND--VTQF----V------- 543
N.haematococca 506 ---KPASS--TQQTPPLVN---VSTPS-TTPT--YSGVPTPPQHD--VTQF----V------- 544
V.albo-atrum 511 ---LLESS--GTADPSARQ---HGN------------------------------------ 527
B.cinerea 528 AP-AVIPN--GLPTPQNTI---IHTPG-SDSN-----LESPLIAD--DEMLY--ENTSL---Y 571
cons 631 . 693

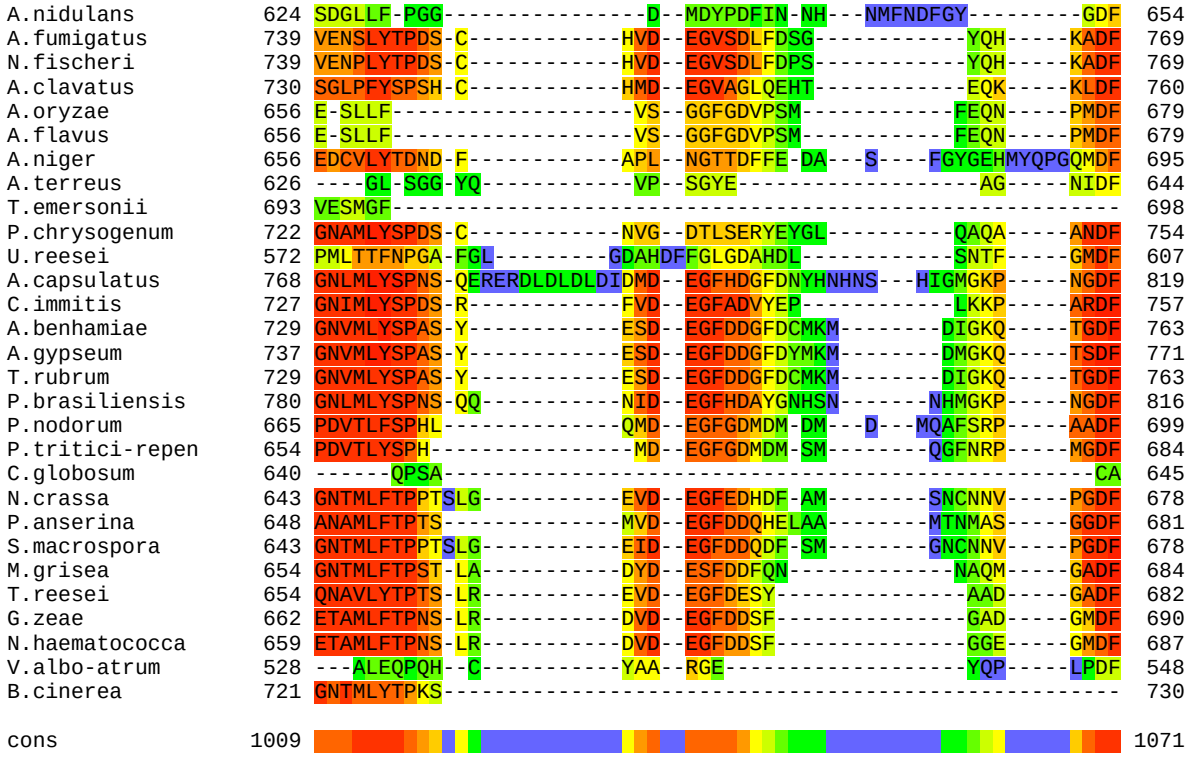

A.nidulans 555 556
A.fumigatus 580 606
N.fischeri 580 606
A.clavatus 572 598
A.oryzae 551 577
A.flavus 551 577
A.niger 548 552
A.terreus 552 578
T.emersonii 571 596
P.chrysogenum 558 584
U.reesei 546 556
A.capsulatus 594 628
C.immitis 570 596
A.benhamiae 549 606
A.gypseum 557 614
T.rubrum 549 606
P.brasiliensis 613 648
P.nodorum 554 571
P.tritici-repen 545 559
C.globosum 545 556
N.crassa 542 552
P.anserina 547 557
S.macrospora 542 552
M.grisea 544 554
T.reesei 540 550
G.zeae 544 554
N.haematococca 545 555
V.albo-atrum 528 527
B.cinerea 572 588
cons 694 756

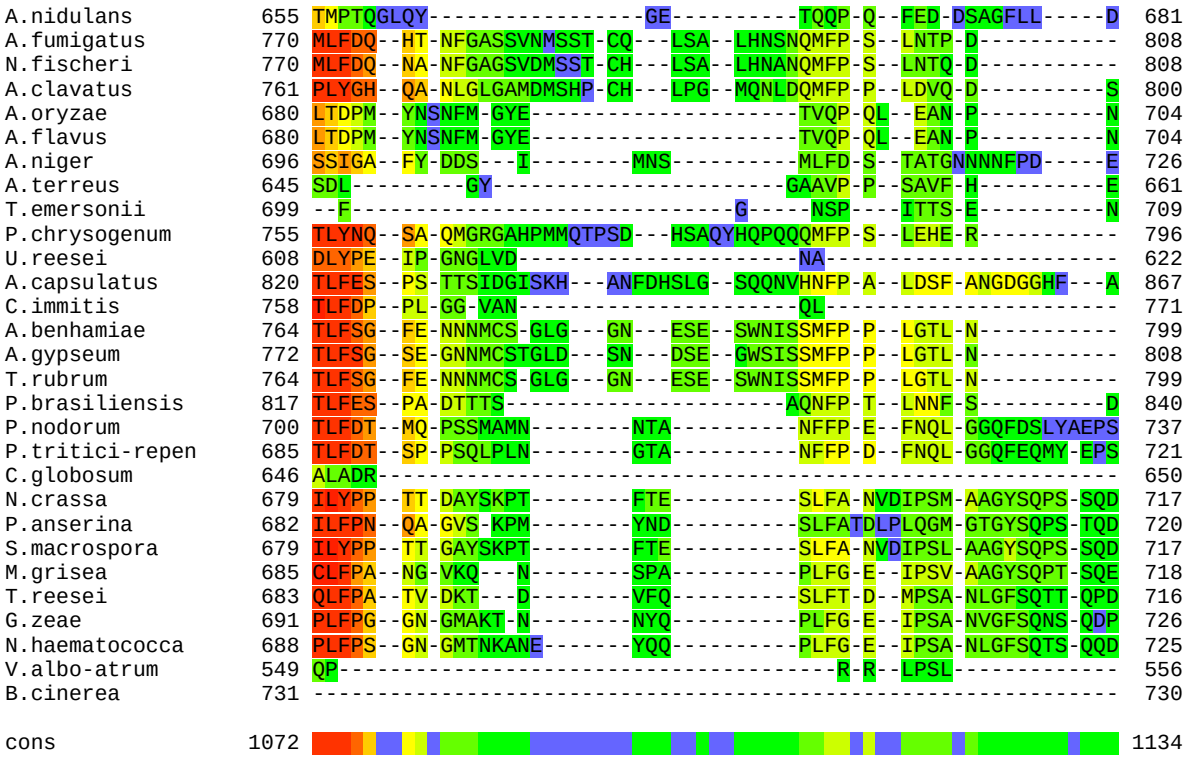

A.nidulans 557 567
A.fumigatus 607 654
N.fischeri 607 654
A.clavatus 599 647
A.oryzae 578 608
A.flavus 578 608
A.niger 553 569
A.terreus 579 580
T.emersonii 597 648
P.chrysogenum 585 627
U.reesei 557 560
A.capsulatus 629 676
C.immitis 597 640
A.benhamiae 607 645
A.gypseum 615 653
T.rubrum 607 645
P.brasiliensis 649 688
P.nodorum 572 600
P.tritici-repen 560 588
C.globosum 557 575
N.crassa 553 569
P.anserina 558 574
S.macrospora 553 569
M.grisea 555 575
T.reesei 551 571
G.zeae 555 574
N.haematococca 556 575
V.albo-atrum 528 527
B.cinerea 589 607
cons 757 819

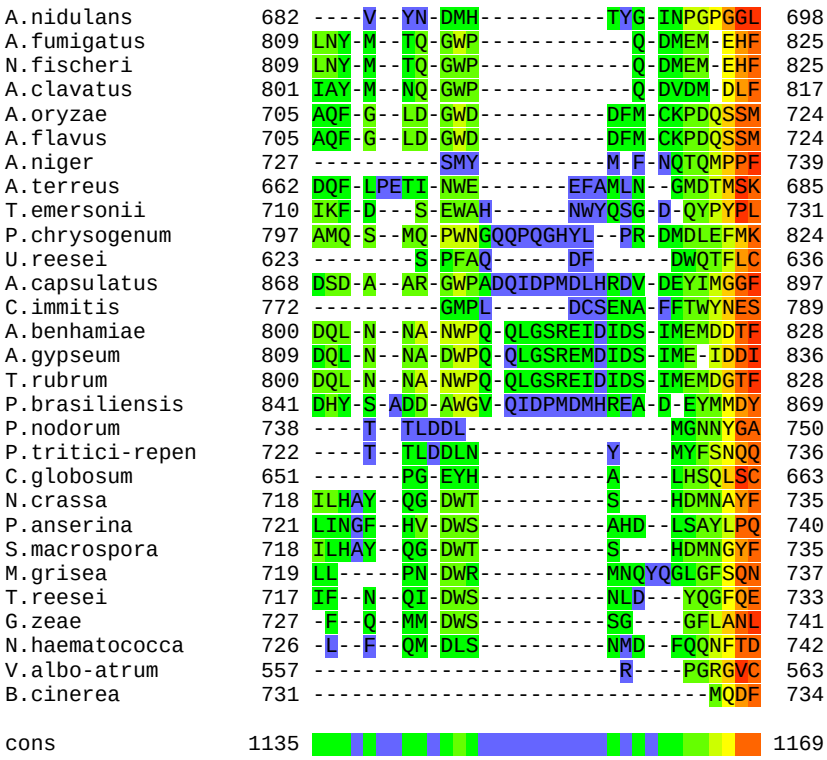

A.nidulans 568 FNTYQAPMTA-MPGSVTG 584
A.fumigatus 655 S-----GSPVP-DLIN--ETMGF----DT----------SPVASTDSSSL--- 682
N.fischeri 655 S-----ASPVP-DLIN--GTLGF----DT----------SPVASTDSSSL--- 682
A.clavatus 648 S-----ASPVP-DLVA--TSMGF----DG----------SPIASTDTSSL--- 675
A.oryzae 609 GI----GSP 613
A.flavus 609 GI----GSP 613
A.niger 570 H-----MSPLQ-L-----S--PD----HMSPLQMTPDRASPAEMLSLQMSP-------D---- 604
A.terreus 581 ----------G-GNGGVNMG-YA----DL----------GFPEMT--RSMDGD 605
T.emersonii 649 SNG---S-HE---SMG--SSSIF----DN----------SPLVMSE-NVD--- 674
P.chrysogenum 628 SMSIPSAESVP-DLMG---PVSF----DG----------SPLTGT--ESI--- 657
U.reesei 561 560
A.capsulatus 677 SSA---TTPDGTGPMG--PGTLT----DE----------SPFEPTH-PTP--- 706
C.immitis 641 SG----ATPIN-EGFG--SA-ML----DQ----------SPMDLSD-SQL--- 667
A.benhamiae 646 ------STPAG--NFA--AG-FQ----AQ----------SPLDLVG-NNA--- 669
A.gypseum 654 ------STPAG--NLA--TG-FH----AQ----------SPPDLIG-SNA--- 677
T.rubrum 646 ------STPAG--NFA--AG-FQ----AQ----------SPLDLVG-SNA--- 669
P.brasiliensis 689 STS---TAPAS-ALLG--SG-LI----DE----------NPFDPRT-TTP--- 716
P.nodorum 601 D-----AVTPE-D-----TA-IN----HN----------DV-------FN--- 617
P.tritici-repen 589 D-----AVTPD-D-----TA-IN----HD----------DV-------YN--- 605
C.globosum 576 N-----PTPST-D-----SG-MD----HT----------SA-------YQ--- 592
N.crassa 570 N-----GTPSP-D-----SG-MD----HN----------SA-------YQ--- 586
P.anserina 575 N-----PTPST-D-----SG-MD----HS----------SA-------YQ--- 591
S.macrospora 570 N-----GTPLT-D-----AG-MD----HN----------SA-------YQ--- 586
M.grisea 576 H-----ATPSS-H-----SN-VSGGSPFQ----------SP-------YQ--- 596
T.reesei 572 N-----LSPAS-A--A--SS-YE----QY----------PP-------YQ--- 589
G.zeae 575 S-----TSPAS-A-----SS-YE----QY----------AP-------YQ--- 591
N.haematococca 576 S-----TSPAS-A-----SP-YE----QY----------AP-------YQ--- 592
V.albo-atrum 528 527
B.cinerea 608 S-----FSPDG-Q-----SP-FD----AN----------SP-------FN--- 624
cons 820 882


A.nidulans 585 TL-----DAVTPTTGTINSPSE----- 601
A.fumigatus 683 NFD-----LAWSQLDAQN--VEEEFTTLTMQMLT--------PEH----- 712
N.fischeri 683 NFD-----LAWSQLDAQN--VEEEFTTLTMQMLT--------PEH----- 712
A.clavatus 676 NFD-----LDWNQLEVQN--LDEEFTTLNMQLLT--------PPS----- 705
A.oryzae 614 A-SGSSEFLT--------PPSG---- 626
A.flavus 614 A-SGSSEFLT--------PPSG---- 626
A.niger 605 QGS-----SV-----QMS-PPQVS-P--AQLSTPATGPLQSPGF----- 634
A.terreus 606 V-----IHM-----NN--FEDMFANVNNPYLD---------- 625
T.emersonii 675 SM 676
P.chrysogenum 658 NFD-----LDWSNLEY-P--ANEDYTAMAAQLPQ--------D-H----- 685
U.reesei 561 TMQLE---------- 565
A.capsulatus 707 NFN-----VDF-----DN--LDNEYTVMNMQLLT--------PAQ----- 731
C.immitis 668 DLN-----IDW-----HD--LENEYTAMNLQLPT--------PE------ 691
A.benhamiae 670 DLD-----FHF-----DV--TENEYTSMNLQLLS--------PAQ----- 694
A.gypseum 678 DLD-----FHF-----DV--TENEYTSMNLQLLS--------PAQ----- 702
T.rubrum 670 DLD-----FHF-----DV--TENEYTSMNLQLLS--------PAQ----- 694
P.brasiliensis 717 NFD-----VDF-----DN--MDSEYITMNMQLLT--------PAQ----- 741
P.nodorum 618 TC-----AL--GSNFT-TGFQQQPT--------PAL----- 637
P.tritici-repen 606 CN-----LL--NFNTI-NTGFQQPT--------PMS----- 625
C.globosum 593 DIG-----TDF-----SM--YDDIY-NARVQLPTP--PPLLTNAY----- 622
N.crassa 587 DLN-----EF-----TL--IDDIY-GATVQLPNQ--VI--SPFY----- 613
P.anserina 592 DIS-----TDF-----TL--YEDIY-SANVQLPTP--MH--ANIY----- 619
S.macrospora 587 DLN-----EF-----TL--IDDIY-GATMQLPNQ--VI--SPFY----- 613
M.grisea 597 DPG-----SDF-----TV--YDDIY-NANAQVQMQ------NPFS----- 622
T.reesei 590 NGS-----TF-----IIN-DEDIY-AAHVQIPAQ--LP--TPEQVY--- 619
G.zeae 592 NGS-----DF-----ILD-NEDLY-AAHMQLPAH--FP--SPEQAVMYS 624
N.haematococca 593 NGS-----AF-----ILD-NEDLY-AAHMQLPTQ--LP--TQDQMAFYN 625
V.albo-atrum 528 527
B.cinerea 625 STSVLGAQDQFQENMQAYPSNDEW-----AAYEQTDLY-SAPAQIHIP--SN--HQIY----- 672
cons 883 945


A.nidulans 602 623
A.fumigatus 713 738
N.fischeri 713 738
A.clavatus 706 729
A.oryzae 627 655
A.flavus 627 655
A.niger 635 655
A.terreus 626 625
T.emersonii 677 692
P.chrysogenum 686 721
U.reesei 566 571
A.capsulatus 732 767
C.immitis 692 726
A.benhamiae 695 728
A.gypseum 703 736
T.rubrum 695 728
P.brasiliensis 742 779
P.nodorum 638 664
P.tritici-repen 626 653
C.globosum 623 639
N.crassa 614 642
P.anserina 620 647
S.macrospora 614 642
M.grisea 623 653
T.reesei 620 653
G.zeae 625 661
N.haematococca 626 658
V.albo-atrum 528 527
B.cinerea 673 720
cons 946 1008


A.nidulans 624 654
A.fumigatus 739 769
N.fischeri 739 769
A.clavatus 730 760
A.oryzae 656 679
A.flavus 656 679
A.niger 656 695
A.terreus 626 644
T.emersonii 693 698
P.chrysogenum 722 754
U.reesei 572 607
A.capsulatus 768 819
C.immitis 727 757
A.benhamiae 729 763
A.gypseum 737 771
T.rubrum 729 763
P.brasiliensis 780 816
P.nodorum 665 699
P.tritici-repen 654 684
C.globosum 640 645
N.crassa 643 678
P.anserina 648 681
S.macrospora 643 678
M.grisea 654 684
T.reesei 654 682
G.zeae 662 690
N.haematococca 659 687
V.albo-atrum 528 548
B.cinerea 721 730
cons 1009 1071


A.nidulans 655 TMPTQGLQY----------GE----------TQQP-Q--FED-DSAGFLL-----D 681
A.fumigatus 770 MLFDQ--HT-NFGASSVNMSST-CQ---LSA--LHNSNQMFP-S--LNTP-D---------- 808
N.fischeri 770 MLFDQ--NA-NFGAGSVDMSST-CH---LSA--LHNANQMFP-S--LNTQ-D---------- 808
A.clavatus 761 PLYGH--QA-NLGLGAMDMSHP-CH---LPG--MQNLDQMFP-P--LDVQ-D---------S 800
A.oryzae 680 LTDPM--YNSNFM-GYE----------TVQP-QL--EAN-P---------N 704
A.flavus 680 LTDPM--YNSNFM-GYE----------TVQP-QL--EAN-P---------N 704
A.niger 696 SSIGA--FY-DDS---I--------MNS----------MLFD-S--TATGNNNNFPD-----E 726
A.terreus 645 SDL---------GY----------GAAVP-P--SAVF-H---------E 661
T.emersonii 699 --F----------G-----NSP----ITTS-E---------N 709
P.chrysogenum 755 TLYNQ--SA-QMGRGAHPMMQTPSD---HSAQYHQPQQQMFP-S--LEHE-R---------- 796
U.reesei 608 DLYPE--IP-GNGLVD----------NA---------- 622
A.capsulatus 820 TLFES--PS-TTSIDGISKH---ANFDHSLG--SQQNVHNFP-A--LDSF-ANGDGGHF---A 867
C.immitis 758 TLFDP--PL-GG-VAN----------QL---------- 771
A.benhamiae 764 TLFSG--FE-NNNMCS-GLG---GN---ESE--SWNISSMFP-P--LGTL-N---------- 799
A.gypseum 772 TLFSG--SE-GNNMCSTGLD---SN---DSE--GWSISSMFP-P--LGTL-N---------- 808
T.rubrum 764 TLFSG--FE-NNNMCS-GLG---GN---ESE--SWNISSMFP-P--LGTL-N---------- 799
P.brasiliensis 817 TLFES--PA-DTTTS----------AQNFP-T--LNNF-S---------D 840
P.nodorum 700 TLFDT--MQ-PSSMAMN--------NTA----------NFFP-E--FNQL-GGQFDSLYAEPS 737
P.tritici-repen 685 TLFDT--SP-PSQLPLN--------GTA----------NFFP-D--FNQL-GGQFEQMY-EPS 721
C.globosum 646 ALADR---------- 650
N.crassa 679 ILYPP--TT-DAYSKPT--------FTE----------SLFA-NVDIPSM-AAGYSQPS-SQD 717
P.anserina 682 ILFPN--QA-GVS-KPM--------YND----------SLFATDLPLQGM-GTGYSQPS-TQD 720
S.macrospora 679 ILYPP--TT-GAYSKPT--------FTE----------SLFA-NVDIPSL-AAGYSQPS-SQD 717
M.grisea 685 CLFPA--NG-VKQ---N--------SPA----------PLFG-E--IPSV-AAGYSQPT-SQE 718
T.reesei 683 QLFPA--TV-DKT---D--------VFQ----------SLFT-D--MPSA-NLGFSQTT-QPD 716
G.zeae 691 PLFPG--GN-GMAKT-N--------NYQ----------PLFG-E--IPSA-NVGFSQNS-QDP 726
N.haematococca 688 PLFPS--GN-GMTNKANE-------YQQ----------PLFG-E--IPSA-NLGFSQTS-QQD 725
V.albo-atrum 549 QP----------R-R--LPSL---------- 556
B.cinerea 731 ---------- 730
cons 1072 1134


A.nidulans 682 ----V--YN-DMH----------TYG-INPGPGGL 698
A.fumigatus 809 LNY-M--TQ-GWP----------Q-DMEM-EHF 825
N.fischeri 809 LNY-M--TQ-GWP----------Q-DMEM-EHF 825
A.clavatus 801 IAY-M--NQ-GWP----------Q-DVDM-DLF 817
A.oryzae 705 AQF-G--LD-GWD----------DFM-CKPDQSSM 724
A.flavus 705 AQF-G--LD-GWD----------DFM-CKPDQSSM 724
A.niger 727 ----------SMY----------M-F-NQTQMPPF 739
A.terreus 662 DQF-LPETI-NWE-------EFAMLN--GMDTMSK 685
T.emersonii 710 IKF-D---S-EWAH------NWYQSG-D-QYPYPL 731
P.chrysogenum 797 AMQ-S--MQ-PWNGQQPQGHYL--PR-DMDLEFMK 824
U.reesei 623 --------S-PFAQ------DF------DWQTFLC 636
A.capsulatus 868 DSD-A--AR-GWPADQIDPMDLHRDV-DEYIMGGF 897
C.immitis 772 ----------GMPL------DCSENA-FFTWYNES 789
A.benhamiae 800 DQL-N--NA-NWPQ-QLGSREIDIDS-IMEMDDTF 828
A.gypseum 809 DQL-N--NA-DWPQ-QLGSREMDIDS-IME-IDDI 836
T.rubrum 800 DQL-N--NA-NWPQ-QLGSREIDIDS-IMEMDGTF 828
P.brasiliensis 841 DHY-S-ADD-AWGV-QIDPMDMHREA-D-EYMMDY 869
P.nodorum 738 ----T--TLDDL----------MGNNYGA 750
P.tritici-repen 722 ----T--TLDDLN----------Y----MYFSNQQ 736
C.globosum 651 --------PG-EYH----------A----LHSQLSC 663
N.crassa 718 ILHAY--QG-DWT----------S----HDMNAYF 735
P.anserina 721 LINGF--HV-DWS----------AHD--LSAYLPQ 740
S.macrospora 718 ILHAY--QG-DWT----------S----HDMNGYF 735
M.grisea 719 LL-----PN-DWR----------MNQYQGLGFSQN 737
T.reesei 717 IF--N--QI-DWS----------NLD---YQGFQE 733
G.zeae 727 -F--Q--MM-DWS----------SG----GFLANL 741
N.haematococca 726 -L--F--QM-DLS----------NMD--FQQNFTD 742
V.albo-atrum 557 ----------R----PGRGVC 563
B.cinerea 731 ----------MQDF 734
cons 1135 1169

Appendix 7 – Conserved secondary structure motifs in StzA proteins.

Secondary structure predictions were obtained for each protein individually using PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/) and then superimposed onto the Advanced TCoffee alignment of all 29 StzA proteins. Helices are highlighted in green and strands in yellow. Coils are represented by residues not highlighted.

```
A.nidulans            1 MSPAQD---SESIKAHPRRRP-FRA-A-R-PSLV--PA-----------EEQS-PSL-PPLR   40
A.fumigatus           1 MSTSQA----EHTRVHPRRRP-LKT--A-PPSLSA-STSIESYVA------DSS-SSS-PNLS   46
N.fischeri            1 MSTSQA----EHTRVHPRRRP-LKT--A-PPSLSA-STSLESYVA------DSS-SSS-PSLS   46
A.clavatus            1 MSAPQA----EPTRIHPRRRP-LKT--A-PPSLIT-SASPE-----------SH-DGP-SSPY   41
A.oryzae              1 MSSAQPV--DESDRRHPRRRV-LNR-P--PPSLSS-DSS------------PPH-LLP-SNLR   42
A.flavus              1 MSSAQPV--DESDRRHPRRRV-LNR-P--PPSLSS-DSS------------PPH-LLP-SNLR   42
A.niger               1 MKS-----------QNPRRGR-HVA-R-P-PSLA--ST------------ADVH-ASR-SRYQ   32
A.terreus             1 MSSPKDSQYDTTLRRHPRRRA-ALS-TS-PPSLLN-SSNCR----------SVD-STP-SNPL   47
T.emersonii           1 MSTQTQ-----QSQVHPRRRP-YKA--G-IPSLST-SSASDQQDAK-----SSS-PCP-PTPR   46
P.chrysogenum         1 MSAIQT----EPSRVHPRRRP-VLN-TA-LPTLSS-D-----------------LPD-SKMP   36
U.reesei              1 MSEP---------QSHPRRRP-FNG-AA-LPSLLT-SSNHGNGPA------SHV-VAP-YKRM   42
A.capsulatus          1 MSAP---------QVHSRRRP-LQH-KN-LPHLSI-TSSSSTSDQSSSDS-TCR-LTP-HSMK   47
C.immitis             1 MSES---------QSHPRRRP-FNK-AAPLPPLLT-SSDHGNGSS------NRA-VRP-HNMK   43
A.benhamiae           1 MSSDL--------SISPRRRP-FLN-AN-LPRLST-GHCDGRSPS------NSS-NTV-RHAT   43
A.gypseum             1 MSSDL--------SISPRRRP-FLN-AN-LPRLST-GHCDGRSPS------NSS-NTV-RHAT   43
T.rubrum              1 MSSDL--------SISPRRRP-FLN-AN-LPRLST-SHCDGRSPS------NSS-NTV-RHAT   43
P.brasiliensis        1 MSAP---------RVHPRRRP-IYN-SAPIPRLPT-SSSPLSSSSDLSDPACRH-SAL-HSMQ   49
P.nodorum             1 MSAPHG------QNSHPRRRP-ING-Q-T----------------------SPL-N-T-HGMS   29
P.tritici-repen       1 MSAPHG------QNCHPRRRP-LKG-T-A----------------------SPL-S-I-HGMQ   29
C.globosum            1 MSS------------NPRRTP-VTR-S-G-S--------------------RGL-SIK-TNVV   25
N.crassa              1 MS-------------NPRRRP-AGG---------------------------L-TLK-TNML   19
P.anserina            1 MSS------------NPRRTP-MTR-P-D-S--------------------RGL-SLK--TNT   24
S.macrospora          1 MS-------------NPRRRP-AGG---------------------------L-TLK-TNVL   19
M.grisea              1 MS-------------NPRRTP-MTR-P-D-S--------------------VL-PLKTTLAV   24
T.reesei              1 MSFS-----------NPRRRTPVTR-P-G-T-----DC-----------EHGL-SLK-TTMT   30
G.zeae                1 MSFS-----------HPRRRTPVTR-P-D-C-----DT-----------ENAL-SFK-NSST   30
N.haematococca        1 MSFS-----------HPRRRTPVTR-P-G-S-----DT-----------ENAL-SLK-TNST   30
V.albo-atrum          1 MSCQ-----------NPRRRSPVTR-V-G-D-----AS-----------SNGLTSLK-TNMT   31
B.cinerea             1 MSFT-----------HPRRTT-KTTSI-A-VTLLANSS-----------SRPT-VNT-TMAS   35
cons                  1 *.              .**                                            63

A.nidulans           41 LRKGETFNPSILRS-SDR-DHL--VPSLPRRSPTCPGALEA--IA-----AGQQRMADILERL   92
A.fumigatus          47 LRKGETFHSPSPPPSDDI-DPVLSFRSLPQRSPTCTRSLE-A-IA-----AREQRMVDYLSNL  101
N.fischeri           47 LKKGETFHSPSPPPSDDL-DPVLSFRSLPQRSPTCPRSLE-A-IA-----AREQRMVDYLSNL  101
A.clavatus           42 LRRAETFHSPSPA-SGDQ-DPVLNFRTLPRRSPTCPRSLE-A-IA-----AGEQRMANILNHL   95
A.oryzae             43 LIKGETFHSSNRPRS-DR-DPILDLKLLPRRSPTCPKALE-A-IA-----AGQRRMAHILNRF   96
A.flavus             43 LIKGETFHSSNRPRS-DR-DPILDLKLLPRRSPTCPKALE-A-IA-----AGQRRMAHILNRF   96
A.niger              33 LKKHQTFHSKSPSSSDDE-DPLADLACSPRRSSTSPAALQA--IL-----AGQQRMNKLLSRF   87
A.terreus            48 FRKSETFHAARNRPT-PR-DPRLSLPLAPRRSPTSPAALE-A-IA-----AGRERMSKILDTL  101
T.emersonii          47 LQKGATFHSPSTPPREDQ-DPVLNIPSLPRRSPTCTKTLE-A-IA-----AGEQRMANILGRF  101
P.chrysogenum        37 LKKGETFHTPTSPPSSDR-DPVLNFRSLPHRSPTS---LE-A-IA-----VAEERMTSILGRL   88
U.reesei             43 LKKGATFHSPTTPSADEC-DPILYIPSLPRRAPTSSRALEEV-IA-----AGERRVAGILGQV   98
A.capsulatus         48 LKKGATFHSPTTPPSEDS-DPIFNIHSLLRRSPTCPKALEDV-IA-----AGVGRRAVFLDKF  103
C.immitis            44 LRKGATFHSPTTPPTEEC-DPILYIPSLPRRAPTCSRALEEV-IA-----AGERRVANILGRV   99
A.benhamiae          44 LRKEATFHVSASPSS-PG-DPVLHLPTLLRRSPTSPDTLQEL-SA-----ARENRVAGWLGSI   98
A.gypseum            44 LRKEATFHVSASPSS-PG-DPVLHLPTLLRRSPTSPDTLQEL-SA-----ARENRVAGWLGSI   98
T.rubrum             44 LRKEATFHVSASPSS-PG-DPVLHLPTLLRRSPTSPDTLQEL-SA-----ARENRVAGWLGSI   98
P.brasiliensis       50 LRKGATFHSPTTPPSPGT-DPILNIHSLLRRSPTCPRSLEDV-IA-----AGEKRMAVFIDKF  105
P.nodorum            30 LRKSDTFSSPKNLSS-DI-CGLENQPFMPRRSPTSPESLVAL-LDQ---HSNVRRITNLLEGL   86
P.tritici-repen      30 LRKSGTFSSPTQISS-DI-CD-LDRHFMPRRSPTNTERLEEL-VQETIQDPSVRRVSNLLKDF   88
C.globosum           26 LKKGATFHSPTTPLE-SP-EPVFTPPSLPRRSHTN---LDDV-ID-----AHRRRVALTLGDI   77
N.crassa             20 LQKGATFHSPTTPASGDSSERVFVPPSLPRRSHTN---LDDV-ID-----SRCRRVALALDAI   73
P.anserina           25 LQKGATFHSPTSPTS-TT-ENVFRPPSLPRRSQSN---LDDV-ID-----SHRRRAALTLDEF   76
S.macrospora         20 LQKGATFHSPNSPASGDSSERVFVPPSLPRRSHTN---LDDV-ID-----SRCRRVALTLDAI   73
M.grisea             25 LTKGSTFS-PISPATPKS-SSSFTPPSLPTRSHTD---LDDV-VD-----AHRRRVALTLGDI   76
T.reesei             31 LRKGATFHSPTSPSASSA-AGDFVPPTLT-RSQSA---FDDV-VD-----ASRRRIAMTLNDI   82
G.zeae               31 LRKGATFHSPTSPSLT-S-DIAFVPPTLP-RAQSH---LDDV-VD-----ANRRRVALTLNDI   81
N.haematococca       31 LRKGATFHSPTSPVSSAS-DLTFVPPRLS-RSQSH---LDDV-VD-----ANRRRIALTLNDI   82
V.albo-atrum         32 LRKGATFHSPTSLDSSSI-D-AFIPPALG-RSQTN---LEDV-VG-----AHVRRMEMIVSGI   82
B.cinerea            36 LRKGATFHSPSSPTN-EV-DGAFNLSALE-RSLSN---LEDFAVE-----SHKRRAADVIESF   87
cons                 64 : :  **                    *: :    :              *    :  .  126
```

```
A.nidulans          93 DLNS-------G------------TT--S-TSDE--NDDLPVPKGLLRLHLQTQARRE-----  126
A.fumigatus        102 NLNSLE--PSS------------PL---SDKDN--GDDLPVPRAILQAHIDSQSMAD-----  139
N.fischeri         102 NLNSLE--PSS------------PL---SDKDN--GDDLPVPRAILQAHIDSQSMAD-----  139
A.clavatus          96 DLDSLD--ATG------------PL---N-SSS--DNDLPVPRGVLQAHFDSEPLAD-----  132
A.oryzae            97 DLDSLS-----------------TR---DSLES--QDELPVPRGILRTHVKSSASKE-----  131
A.flavus            97 DLDSLS-----------------TR---DSLES--QDELPVPRGILRTHVKSSASKE-----  131
A.niger             88 DLDS-------S------------DS--A-PRKE--HDGLPVPRGYLEIHFERSQSGK-----  121
A.terreus          102 DLDTFT-----------------P----SESV--DEELPVPRSVLQLHFDSLHISD-----  134
T.emersonii        102 DFDSLS--PS--------------KA---TNRQE--EDDLPVPRGILQAHVGQQ-RMQ-----  137
P.chrysogenum       89 TLEPTQ--DQS------------S------EVA--GGESPLDSGLGRNNSKSMPDSP-----  123
U.reesei            99 ERDLAGIDGKST-----------RN---RPSFP--GDDFPVPRGLLHARVADTDPMD-----  139
A.capsulatus       104 ERSPYGLGSDLS-----------APRL-DNTLC--DNDSPVPRGMPEARMTNTDPMN-----  146
C.immitis          100 ERDLAGRDGEPS-----------RS---RPSFP--GDDSPVPRGILAAHIANTDAMD-----  140
A.benhamiae         99 ERNSTGPGNKSL-----------SK---QSSLEDDKRNLKALRGLSK-RLTSANLME-----  140
A.gypseum           99 ERNSTGPGNKSL-----------SK---QSSLEDVNSSLKAPRGLNT-RLTSANLME-----  140
T.rubrum            99 ERNSTGLGNKSL-----------SK---QSSLEDDKRNLKALRGLSK-RLTSANLME-----  140
P.brasiliensis     106 ERNLSGLGTCSPSTASSSSSSSSSPRNRHSILD--DDDLPVPRAILDAHVIHADTMD-----  161
P.nodorum           87 DKQLSGHK-AAS----------AIA--ANILAD--PEVLPVPSFVLDNATLDTTPME-----  128
P.tritici-repen     89 EARVAGHK-TSN----------T-G--ASILSD--PDVFPVPSLFLDNTSMDCTPME-----  129
C.globosum          78 DKTLAGLRNDDA----------PLSP-GRKTLR--DEILPLPRGVLDHTLDSVMV-K----E  121
N.crassa            74 ERQLASS--NDT----------FAS--A--SRS--DKCIPPPRGLLERNLDSPIMPK----E  113
P.anserina          77 DRTLAGLSISDS----------PSSAAARKILR--EDSPPIPRGILNHTLDTVMA-K----G  121
S.macrospora        74 TRQLASS--NDT----------FVS--A--KQS--GECIPPPRGLLQRNLDSPIMPQ----E  113
M.grisea            77 EKTLAGMS-LDS----------PSA--N-KAFR--DNSYPLPRGLLDPTLIGKEM------D  116
T.reesei            83 DEALSKASLSDK-----------SP--RPKPLR--DTSLPVPRGFLEPPVVDPAMNK----Q  125
G.zeae              82 DEALAKTQELSLSS----------TS--KPMTLR--DTGLPIPRGFLEGPIVDPKMTK----E  126
N.haematococca      83 DEALAKTEELSLSP-------RSPKS--KPKSLR--DTSLPIPRGFLEGPIVDPAMAK----E  130
V.albo-atrum        83 ETSLNLND--TP------------RP--ASKPSR--DECLPRTNGFLGRPTVDPAMAK----D  123
B.cinerea           88 ELTVAGN--QSP------------ST--PRRRFR--DESDPVPRGVVNVENPSTKAYNTELMD  132
cons               127                                                               189

A.nidulans         127 -GT--VE-PHSRQPSPMPK--EHSR----K-AQR-VHCHASDSGIGSSISSAQSVSSNKVKA-  176
A.fumigatus        140 -RV--SR-PT-QSSE---L--HTPS----K-VRK-THCHASDSGLGTSLSSENMSASDK--MK  184
N.fischeri         140 -RV--SK-PA-QSSQ---L--HTPS----K-VRK-THCHASDSGLGTSLSSENMSASDK--MK  184
A.clavatus         133 -KP--TK-RP-HSPQ---M--EPPR----K-VRK-SHSHASDSGLGTSLGSEALSASNKSKVK  179
A.oryzae           132 -DS--TK-Q--SEPT---P--QEPKESHKK-IRR-VNHHTSDSGLGSSIGSAETMSSTKG---  178
A.flavus           132 -DS--TK-Q--SEPT---P--QEPKESHKK-IRR-VNHHTSDSGLGSSIGSAETMSSTKG---  178
A.niger            122 -GD--SF-HYDLRRGSSRK--DSQT----K-VRK-VHCHPSDSGLGTSISSCETASARQAKE-  171
A.terreus          135 -PP--QS-P--S-VR---S--APKK----P-LQR-VNHHTSDSGLGTSICSEEVLSCT-----  174
T.emersonii        138 -TP--RK-EH-SLPT---PPADTPN----K-VQTAHHRHASDSGLGSSISSTAANERVKAGHL  187
P.chrysogenum      124 -HS--SN-ED-GRSI---PIPLDEK----K-ARQ-QHSHESDSGLGTSVSSEEGLADSK--NK  170
U.reesei           140 -ID--TA-PYDSKFH---SSR-RLP----P-VDA-KEHRVADSGIGSSILDAPLEKTP-----  183
A.capsulatus       147 DVD--SI-QKDKAFK---PV----R----D-QDR-RRRLSCDSGIGSSISGASMSSSVDTREQ  193
C.immitis          141 -VD--TA-TNHARPN---RSR-RLP----P-VDI-KHRRTSDSGIGSSIGEALPHKSAAGSGQ  189
A.benhamiae        141 -ID--SS-FR-NQVQ---TG---PP----G-KEN-LRAELSDSGVGSSVGEDILKRE------  180
A.gypseum          141 -ID--SS-FR-NSTE---SG---FP----G-KEN-IRTELSDSGVGSSIGEDVLNRERAGAVE  186
T.rubrum           141 -ID--SS-FR-NQVQ---TG---PP----G-KEN-LRAELSDSGVGSSVGEDILKRE------  180
P.brasiliensis     162 -TD--ST-DS-NHCK---PLRIDDS----P-AMN-SSQHASDSGLGTSVSGTSNSASIHSREK  210
P.nodorum          129 -----LD-----------S--KPSA----V-QRN-THEHASDSGLGSSIGGSKYGKHAARTQ-  166
P.tritici-repen    130 -----LD-----------V--DVFK----S-AAA-EHDHASDSGLGTSIGESKYDDIRTQ-R-  166
C.globosum         122 -QP--AE-R--R------M--LRPR----T-RRS-SRFHESDSGLGTSIASTNDKDAVAKKD-  163
N.crassa           114 -VE--PE-R--R------M--LRPR----T-RRS-SRHHDSDSGLGSSIASTSEKDAS-SKA-  154
P.anserina         122 -KE--VE-R--K------V--LRPR----T-RRT-SRHHDSDSGLGTSIASTNEKIAA-KEQ-  162
S.macrospora       114 -AE--NE-R--R------M--LRPR----T-RRS-SRHHDSDSGLGSSIASTSEKDAS-CKA-  154
M.grisea           117 -----VE-R--R------V--LRPR-PVRR-PGQ-SHHHESDSGLGSSILSTKQKSTP-NAD-  158
T.reesei           126 ----EPE-R--R------V--LRPR----S-VRR-TRNHASDSGIGSSVVSTNDKAGAADST-  166
G.zeae             127 -----EE-R--R------T--LRPR----GRTSR-ALEDHSDSGLGTSVASTNEKRGAVTAS-  167
N.haematococca     131 -----AE-R--R------V--LRPR----S-VRR-SRHHASDSGLGTSVASTNEKHAAVTQS-  170
V.albo-atrum       124 -TKTSGE-R--R------V--LRPR----H-RRS-SEQHASDSGLGTSLASSVEKQAPSITS-  167
B.cinerea          133 GVM-PSEQS--S------V--RRST----R-PRR-TTPRYADSGLGSSIGSSSSDKKVLTND-  177
cons               190                                        .***:*:*:              252

A.nidulans         177 -GQ--------L-SRSNLPT--SRSQSAITRSISAM-DAQSTQRHKLSSEGRAEIEKHVIGPL  226
A.fumigatus        185 AGQ-LSFES---QSSAKATT--RMTQSAITSSISAFD-AKISQKRQLGLPACKQIERLIIKPI  240
N.fischeri         185 AGQ-LSFES---QSSAKAAT--RMTQSAITSSISTFD-AKVSQKRQLGLPACKQIERLVIKPI  240
A.clavatus         180 AGQ-LSFEP---QVSGMSGS--QMTQSAITSSISAIN-PKQIAKYQLGARACKQIEKNLLVPI  235
A.oryzae           179 ---------------------------NVTAGQVCQ----ASVGHKLSATARVEIQGRILFPL  210
A.flavus           179 ---------------------------NVTAGQVCQ----ASVGHKLSATARVEIQGRILFPL  210
A.niger            172 -------------SR-SKTV--TRTQSAITSSLMAV-EPDSASKQNLSLGASQEIEKRILTPL  217
A.terreus          175 ------------------------ESADLDSNAPM-S--ASHLRKISPDGVDKIEHRVFYPL  209
T.emersonii        188 -C------------TGNADL--PGGQSAITHSISSGD----TSRRQLPLAACKQIERFILAPI  231
P.chrysogenum      171 V---------------NDGI--HDNQSAITSSISAFD-TGSSPKRQLGPAACKQIERFVLVPI  215
U.reesei           184 ------------------------------------ASGGPLMRFGAFKQIERHILAPI  206
A.capsulatus       194 VRK-ATHSF-----QEEGKI--VSTQSAITRSISSTT-AL-AQIPSFSVTARKHIDRFILVPI  246
C.immitis          190 ------------------PRK--SPSRSAITRSISYKSSTMDASKPVLSPTAVKQIERRILIPI  233
A.benhamiae        181 ----------------------------AR--ESVV----SSSSTLSSGGKRHIQRHILLPL  208
A.gypseum          187 ---------------------------VSTR--ESVA----SSRTKLSTGAKRHIQRHVLLPL  216
T.rubrum           181 ----------------------------AR--ESVV----SSSSTLSSGGKRHIQRHILLPL  208
P.brasiliensis     211 AIDKMNHSPKDRKSLQEPKM--ASTQSAITRSISSNT----LEIPTFGVAARKNIDRFILNPI  267
P.nodorum          167 PSS-------VS-DSVNTSI--TSTHSAITRSMSSL-GA-SEEKHTLGEYACQQIHDNIIKPI  217
P.tritici-repen    167 -SS--------R-ESISSSI--SSTHSAITRSFAAL-GA-SDEDHTLGEYACKQIHDSIIKPI  215
C.globosum         164 ------------------K---AGRTTAVTRSAAAR-SATAASAVGLSPRATSRIYAHTLKPL  204
N.crassa           155 ------------------K---TTRTSAVARSATAR-AASTPDLPGLGDRATNRIVEYILKPL  195
P.anserina         163 ------------------T---VAKTTAVTRSAAATRTTTTTTTQVLGQRANNRICEHTLKPL  204
S.macrospora       155 ------------------K---TTRTSAVTRSAAAR-SASALNLPGLSDRATNRIVEYVLKPL  195
M.grisea           159 --------------STPAK--TGVKGSAVTRSAAS---STSKNLPSLSARATNRVFEHTLKPL  202
T.reesei           167 ------------------KK---PQASALTRSAASS-T--TAMLPSLSHRAVNRIREHTLRPL  205
G.zeae             168 ------------------KE-AKVQTRCLTRSAAAA-AA-TGKLPSLGSKAFSRIHEHTLRPL  209
N.haematococca     171 ------------------KE-T-KTRSAITRSAAAS-S--AEKLPSLGSKAINRIHEHTLRPL  210
V.albo-atrum       168 ------------------KT---SKASAITRSAAAP-SNTMTKVSGLSSKAVSRVHEHVLRPL  208
B.cinerea          178 ------------------NVSKSTKPSTITRSAAAP-STTIKSLPRLSARTTARIQEHILKPL  221
cons               253                                                :      .:    : *:  315
```

```
A.nidulans          227 LEDEKSKPFHPILEDVRQQIDDERISCLRDLEKTVFSLAP-EVKTNDAAYVRFCQYTILCLGQ  288
A.fumigatus         241 LREERLKPFHPLVQSIPQRIVDREIGCLRDLEKTLLWLAP-NYAASRASYLRFCEFTIQCLHT  302
N.fischeri          241 LKEERLKPFHPLVQSIPQRIVDREIGCLRDLEKTLLWLAP-NYAASRASYLRFCEFTIQCLHT  302
A.clavatus          236 LGEEKLKAFHPLVRSVPQRIVSKEINCLRDLEKTLLWLAP-HYATTRASYIGFCEFTIQCLYS  297
A.oryzae            211 LMKDKFQLFHPLVRCAHQQIEKQQLKCLRDVEKTLLFSTP-DVKAPTAAWYSFCRFTIHCLHE  272
A.flavus            211 LMKDKFQLFHPLVRCAHQQIEKQQLKCLRDVEKTLLFSTP-DVKAPTAAWYSFCRFTIHCLHE  272
A.niger             218 LGQELCKPFHHLVESAREQIEKKQIGTLRDLEKSLLHGAA-DVEAEVGSYYQFCSSTILCLSE  279
A.terreus           210 LMNTQFDKFHSTVRYAAQGVKDNRFRCLRDVENLFRNPFPIDHVNEKEEFSLFLNQVVLCLDD  272
T.emersonii         232 LREERLKPFHPLVRSIPQRVVDKEITCLRDLEKTLLWLAP-QYSITKASYLGFCEFTIQSIHT  293
P.chrysogenum       216 LKEPRLKPFHPLVRSVPQRIINKQIVCLRDLEKTLLWLAP-NSATSRNSYLNFCEFTIQCLHT  277
U.reesei            207 LGEQSFKPFHPIVETIPQRVRNNNITCLRDLEKILLFVPL-NETISKGSYMQYGRFTIHCLHT  268
A.capsulatus        247 LREKRLSPFHPLIRSVPQRIESKEIACLRDLEKTILFLAP-RYTQSKAVYLGFCEFTIQCLHT  308
C.immitis           234 LREKHLGHFHPLVRGVPQRIKLNDITCLRDLEKTLLFLAP-DFAPTRTSYISFCDFTIQCLHT  295
A.benhamiae         209 LKEQRLKNFHSLVRSVPQRIQSSEITCLRDVEKTLLYLAP-HHTISRSSYLSFCEFTIQCVHT  270
A.gypseum           217 LKEQRLKTFHSLVRSVPQRIQSSEITCLRDVEKTLLYLAP-HHTVSRSSYLSFCEFTIQCVHT  278
T.rubrum            209 LKEQRLKNFHSLVRSVPQRIQSSEITCLRDVEKTLLYLAP-HHTISRSSYLSFCEFTIQCVHT  270
P.brasiliensis      268 LLEKRLTPFHPLARGVPERIENKEITCLRDLEKTLLFLAP-KFTTSKASYIAFCEFTIQCLHT  329
P.nodorum           218 LAEESLKDFHPLIQDVPRRIGEKSICNLRDLEKTLIFLAP-ELSASATSYLQFCELSIQLLHA  279
P.tritici-repen     216 LAEDSLKDFHSLVKDVPRRIGEKNIRNLRDLEKTLIFLAP-ELSATSDSYSRFCQRSIQLLAT  277
C.globosum          205 LANSGLKDFHPLLLECPRKIQERQIVCLRDLEKTLLLVAP-ERTKSSELYLDFCLTTIRCIQA  266
N.crassa            196 LAKPNLKEFHSLVLECPKKIQEKEILCLRDLEKTLILVAP-ERTKAAGSYLDFCLTTIDCLQA  257
P.anserina          205 LGKPEFKDFHPLLLECPKKIQDKEIVCLRDLEKTLLLVAP-ERTKSAGLYLDFCLTTIQCIQA  266
S.macrospora        196 LAKPALKEFHSLVLECPKKIQDKEILCLRDLEKTLILVAP-ERTKAAGSYLDFCLTTIDCLQA  257
M.grisea            203 LADSSFKEFHPVLLECPRKIQSKQIVCLRDVEKTLLLTAQ-GVTKVPKLYLDFCLSTVQCIRA  264
T.reesei            206 LEKPTLKEFEPIVLDVPRRIRSKEIICLRDLEKTLIFMAP-EKAKSAALYLDFCLTSVRCIQA  267
G.zeae              210 LAKPTLKEFKPIVLDIPRRIQSKEIICLRDLEKTLIFMAP-EKAKSATLYLDFCLTSVRCIQA  271
N.haematococca      211 LAKPTLKDFEPIVLDIPRRIRSKEIICLRDLEKTLIFMAP-ERTKSAALYLDFCLTSVRCIQA  272
V.albo-atrum        209 RAKPELKDFEPIVLDIPRRIRDKEIICLRDLEKTLIFMAP-ERAKTAALYLDFCLTSIRCIQA  270
B.cinerea           222 LAKESFKDLHPLVQDCPRRIHQKEIVSLRDLEKTLIFTAP-KYTKSASLYLDFCLESIHCIQA  283
cons                316   .     :.       . :    :  ***:*: .              :  :    :  :   378

A.nidulans          289 TVSFLNGRDLCLPTDKQYNNGYFVDLLDQVSQFKRIRDEWKRRHE--------A-D---GKVK  339
A.fumigatus         303 SVYHLNERDQRLPADRPYTNGYFLDLVAQIRRYAAMINESRSSMPSN---REPAQNGAKASAP  362
N.fischeri          303 SVYHLNERDQRLPADRPYTNGYFLDLVAQIRRYAAMINESRSSMPSN---REPAQDGAKASAP  362
A.clavatus          298 TVQHLNSRDQRLPTDRPYTNGYFLDLVAQIHQYAAMINASRNQMSSN---E-KP-AR-ASASS  354
A.oryzae            273 TSGYLRGRDLTLPNDVPYSNNYFLDLITQINRFARIRDATRSRQES----ATNG-EK-AAKSL  329
A.flavus            273 TSGYLRGRDLTLPNDVPYSNNYFLDLITQINRFARIRDATRSRQES----ATNG-EK-AAKSL  329
A.niger             280 TYTHLDPRDLCIPTDKSYSNSYFLDLTAQVHRFKAMRDEARKNKK---------------EAS  327
A.terreus           273 TWHRLDESDRTMPGDVPYSNEYILDLYDQIVRFKALCERAKQNLS------QKD-SN-GATSK  327
T.emersonii         294 AVTHLNERDQRLPTDRPYTNGYFLDLVAQIRQYAAMIAASRERARAR---GEQA-DA-ARHNS  351
P.chrysogenum       278 SASHLNDRDQRLPADRPYTNGYFLDLVSQVRRYAAMVRAERERVQAT---QGSESNKETKSPI  337
U.reesei            269 AVRHLNDRDQCLPADRPYTNGYFLDLVAQVQGYAATIAAARNSRSSR---DQKG-NN-LDYSS  326
A.capsulatus        309 TVGYLNDRDQRRPSDRPYTNGYFLDLVEQMRQYAALVGATRERQGPRPRSRIAE-NN-MEYSS  369
C.immitis           296 TVGFLNDRDQRRPADRPYTNGYFLDLVEQVRQHAVMLHSTRQSQG------KNE-NN-LDYSS  350
A.benhamiae         271 TVGYLNDHDQRRPADRPYTNRYFLDLVEQIRQYAALIRSSRERRAT----SSNR-DE-LDYSS  327
A.gypseum           279 TVGYLNDHDQRRPADRPYTNGYFLDLVEQVRQYAVLIRTARERRAA----SPNR-DD-LGYSS  335
T.rubrum            271 TVGYLNDHDQRRPADRPYTNGYFLDLVEQIRQYAALIRSSRERRAT----ASNS-DE-LDYSS  327
P.brasiliensis      330 TVGYLNERDQRRPSDRPYTNGYFLDLVEQIRQYASMISAARERQGPRPRSRDGE-NN-KDYSS  390
P.nodorum           280 TVEQLSEQDQRLPSDRPYTNNYFLDLAEQIQRYARIMAATRQKESK----GESL-DD-MDYSR  336
P.tritici-repen     278 TVQYLSEQDQRLPSDRPYTDGYFLDLMEQIRRYASIMAATREKEVK----GEDL-DD-MDYSR  334
C.globosum          267 TVEYLSDREQTRPRDVPYTSGYFVDLVDQIRNYAQQLADAKENQ-----------DE-MDVDP  317
N.crassa            258 TVQHLGDRELTRPRDLPYTSGYFVDLVDQFYNYARQIAESNKTKE-------GA-ND-MDIDP  311
P.anserina          267 TVEYLSDREQTRPRDVPYSSGYFIDLVDQIRHYAQQLSEAKEKG--------EN-DE-MDVDP  319
S.macrospora        258 TVQHLGDRELTRPRDLPYTNGYFVDLVDQFYNYARQIAESNKTKD-------GA-NG-MEVDP  311
M.grisea            265 TVVYLSETEQRRPKDVPYSSGYFIDLVDQIKQYAQQLADQKKAG-----------A-SDPNI  314
T.reesei            268 TVEYLTDREQVRPGDRPYTNGYFIDLKEQIYQYGKQLAAIKEKGS-------LA-DD-MDIDP  321
G.zeae              272 TVEYLSDREQIRPADRPYTNGYFLDLKDQILEYGKQLAAKNS----------G-DE-MDIDA  321
N.haematococca      273 TVEYLSDREQIRPADRPYTNGYFIDLKDQILEYGKQLAAKNN----------G-DD-MDVDA  322
V.albo-atrum        271 TVEYLSDREQIRPADRPYTNGYFIDLVEQIRQYAGQLATAKEAGV-------EG-RE-MDVDP  324
B.cinerea           284 TVELVNEREQTRPNDRPYTNGYFVDLVEQIKQYAEQVREAKEKEEK----GETI-SE-MDAHP  340
cons                379 :   :   :   * *  *.. *::**  *.  .       .                     441

A.nidulans          340 APQLRLEGGLSQTGRLLEMVV--EQDGEAISLRTGKPYEG-----QPIPSMKRSLSAAS-TD-  393
A.fumigatus         363 NEHITLEGGLSKNGRPAELVVH--KDGEMISLRTGKPYEENA-----VPAMKRSLSLGS-VD-  416
N.fischeri          363 LEHVTLEGGLSKNGRPAELVVH--KDGEMISLRTGKPYEEDA-----VPAMKRSLSLRS-VD-  416
A.clavatus          355 SDPVTLEGGLSKNGRPAELVVH--KDGEMISLRTGKPYEEGA-----TPSIKRSLSIGS-VD-  408
A.oryzae            330 EPRLTLEGGMSETGRPAELVMH--KDGKSISLQTGKPYDEH-----AIPTFKRTLSVET-VD-  383
A.flavus            330 EPRLTLEGGMSETGRPAELVMH--KDGKSISLQTGKPYDEH-----AIPTFKRTLSVET-VD-  383
A.niger             328 ELKLTLEGGMSQTGRPLELVA--QKEGQAVSMQTGEAYDA-----HAPPLVKRTLSLNE-AD-  381
A.terreus           328 LPKLVFKGGLAKTGRQAELVAQ--SDDQMMSLRTGKPYQES-----QVPSMKRGLSLCS-TDG  382
T.emersonii         352 EERLTLEGGLSQTGRPAELVCH--KDGKAISLRTGEPYDEKA-----VPTMKRTLSNQS-ID-  405
P.chrysogenum       338 IPKATLEGGLSQNGKSAELVVM--QDGKPISMATGKPYEADI-----PGVAKRTISLDDAVD-  392
U.reesei            327 GEELVLEGGLSSSGRPAELVRR--KKDVAISLQTGEPYVESK---PTIPIMKRALSVEA-GD-  382
A.capsulatus        370 DEEVTLEGGLGATGRPAELVRR-KKDKQAISLRTGNPYDEKA--SPAVPIMKRSLSMES-CD-  427
C.immitis           351 GEELHLEGGLSSTGRPAELVRR--KNGVAISLKTGEPYVESK---PIMPAMKRSLSIDS-GD-  406
A.benhamiae         328 GEELVLEGGLASTGRPAELVRK--KKGQAISLKTGEPFVEQK---PEVPSLKRAPSMEN--Q-  382
A.gypseum           336 GEELVLEGGLASTGRPAELVRK--KKGQAISLKTGEPFVEQK---PVVPSLKRAPSMEN--Q-  390
T.rubrum            328 GEELVLEGGFARTGRPAELVRK--KKGQAISLKTGEPFVEQK---PEVPSLKRAPSMEN--R-  382
P.brasiliensis      391 DEELTLEGGLAATGRPAELVRR--KKNQAISLRTREPYEERA--SPVIPMMKRSLSTES-CD-  447
P.nodorum           337 DEKITLRGGFSHDGRPMQLVR--EKNGKSMSLAEGIDS--------SYKVSKRPLSDDE-MDE  388
P.tritici-repen     335 DEKITLRGGLSHNGRPVELVR--EKNGKVVPLADS------------PTIKRRFSDED-EDE  381
C.globosum          318 TDEIKLHGGIHINGRPAELVRV-KKNGKMISMATGQPIDKIDEEADGTVRFKRSASQEL-EDE  378
N.crassa            312 TDQIKIHGGPHINGRLSELVRV-KKNGQAISLATGLPVDLCDKAPETPVNFKRSQSEEA-LDE  372
P.anserina          320 TDEIKLHGGIHINGRPAELVRI-KKNGKMISMATGEPIESIEEESSGAVRIKRSASEEL-EDE  380
S.macrospora        312 TDQIKIHGGPHINGRLSELVRV-KKNGQAISLATGLPVDLCEKAPESPVNFKRSQSEEA-LDE  372
M.grisea            315 KEEVKLHGGIAVNGRPAELVR--ISGGKAISLATGEPVELDE-ETTSPIRFKRSASQQL-ADD  373
T.reesei            322 SDEVRLYGGVAENGRPAELIRV-KKDGTAYSMATGKIVDMT----ESPTPLKRSLSEQR-EDE  378
G.zeae              322 SDEIKLVGGLSVNGRPAELVRV-RKDGTYISLDTGKPVETDD---DAPMKMKRSLSQQL-EDE  379
N.haematococca      323 SDEIKLIGGIAENGRPAELVRV-RKDGTYISLETGEPVEVD----EAPMKMKRSLSQQL-EDE  379
V.albo-atrum        325 TDEVKLFGGISQNGRPAELVRV-RKDGQAISMATGLPVDMDE-DGKDFPRLKRSLSQQL-ADD  384
B.cinerea           341 SDEVKLHGGLTRNGRPAELVRINKKTGKAISIATGQPVDMDD-DDNKSMRFKRSLSEEA-EDE  401
cons                442     : **    *:  :::    .    .:                  **  *        504
```

```
A.nidulans          394 EGVQRSMARRKKNAPPMNIN-KKCK--DCDKVFARPCDLTKHEKSHSRPFKCPVTSCKYHIKG  453
A.fumigatus         417 EGVERSMARRKKNAPPMDIN-QKCK--DCDKVFKRPCDLTKHEKTHSRPWKCTEPSCKYHEIG  476
N.fischeri          417 EGVERSMARRKKNAPPMDIN-QKCK--DCDKVFKRPCDLTKHEKTHSRPWKCTEPSCKYHEIG  476
A.clavatus          409 EGVERSMARRKKNAPPMDIN-QKCK--DCDKVFKRPCDLTKHEKTHSRPWKCNDDSCKYFEVG  468
A.oryzae            384 EGVERSMARRKKNAPPMDIN-QKCQ--FCDKVLKRPCDLTKHEKTHSRPYKCPERGCKYFELG  443
A.flavus            384 EGVERSMARRKKNAPPMDIN-QKCQ--FCDKVLKRPCDLTKHEKTHSRPYKCPERGCKYFELG  443
A.niger             382 EGARRSMARRKKNAPPMNIN-TKCS--HCDKIFQRPCDLTKHEKTHSRPFKCPFEGCKYHELG  441
A.terreus           383 EGAHRCMARRKNEAPMNIN-TPCD--FCGQIFARPCDLTKHMKTHTRPFKCSVPECKYYTYG  442
T.emersonii         406 EGVARSMARRKKNAPPMDIN-QKCR--DCDKVFKRPCDLTKHEKTHSRPWKCTEPTCKYFEIG  465
P.chrysogenum       393 EGVERSMARRKKDAPPMNIN-QKCA--DCDKVFKRPCDLTKHEKTHTRPWKCEYTDCSYHTKG  452
U.reesei            383 DSVMRSMARRKKNEPPLNIN-KKCD--HCDRIFRRPCDLSKHEKTHTRPWKCTEPGCKYSKTG  442
A.capsulatus        428 DGVARSMARRKKNAPPLDIN-QKCK--DCDKVFKRPCDLTKHEKTHSRPWKCAEKSCKYFEIG  487
C.immitis           407 DGVMRSMARRKKNEPPLNIN-KKCD--HCEKVFRRPCDLTKHEKTHTRPWKCTEPTCKYSKVG  466
A.benhamiae         383 ESVMRSMARRKKDAPPMNIN-EKCK--HCDKVFRRPCDLTKHEKVHTRPWKCLDTKCKYYQLG  442
A.gypseum           391 ESVMRSMARRKKDAPPMNIN-EKCK--HCDKVFRRPCDLTKHEKVHTRPWKCLDTKCKYHQLG  450
T.rubrum            383 ESVMRSMARRKKDAPPMNIN-EKCK--HCDKVFRRPCDLTKHEKVHTRPWKCLDTKCKYYQLG  442
P.brasiliensis      448 DSVTRSMARRKKNAPPLDIN-QKCK--DCDKVFKRPCDLTKHEKTHSRPWKCDERSCKYFEIG  507
P.nodorum           389 DEATRSMARRRKSDKPGDVM-HTCR--DCKKEFKRPCDLTKHEKTHSRPWKCSEEKCKYFDLG  448
P.tritici-repen     382 DEVTRSMARRRKSEKPGDVM-HTCR--DCKKQFKRPCDLTKHEKTHSRPWKCSEKNCKYYDLG  441
C.globosum          379 EEIMRSMARRKKNASPEELAPKKCREPGCNKEFKRPCDLTKHEKTHSRPWKCPVPTCKYHEYG  441
N.crassa            373 EEVMRSMARRKKNASPEELAPKRCREPGCNKEFKRPCDLTKHEKTHSRPWKCPVKTCKYHEYG  435
P.anserina          381 EEIMRSMARRKKNATPEELAPKKCREHGCNKEFKRPCDLTKHEKTHSRPWKCPVTTCKYHEYG  443
S.macrospora        373 EEVMRSMARRKKNASPEELAPKRCREPGCNKEFKRPCDLTKHEKTHSRPWKCPVKTCKYHEYG  435
M.grisea            374 DEIMRSMARRKKNAPPEEYAPKMCREPGCGKEFKRPCDLTKHEKTHSRPWKCPVPTCKYHEYG  436
T.reesei            379 EEIMRSMARRKKNATPEDVAPKKCREPGCTKEFKRPCDLTKHEKTHSRPWKCPIPTCKYHEYG  441
G.zeae              380 EEIQRSMARRKKNASPEELAPKKCREPGCNKEFKRPCDLTKHEKTHSRPWKCPVSTCKYHTYG  442
N.haematococca      380 EEIQRSMARRKKNASPEELAPKRCREPGCNKEFKRPCDLTKHEKTHSRPWKCPVPTCKYHTYG  442
V.albo-atrum        385 EEIMRSMARRKKNATPEELAPKKCREPGCNKEFKRPCDLTKHEKTHSRPWKCPVKTCKYHEYG  447
B.cinerea           402 ESVMRSMARRKKDASAAELAPKFCREKGCDKSFKRPCDLTKHEKTHSRPWKCPVTSCKYHEYG  464
cons                505 :   *.****:*.  . :     *    * : : *****:** * *:**:**    *.*   *  567

A.nidulans          454 WATEKESERHYNDKHSDAPRLFACQFESCSYKSKRESNCKQHMEKTHGWVYMRSKNNGRSKAS  516
A.fumigatus         477 WPTEKERDRHINDKHSKAPALYKCKFAPCTYSSKRESNCKQHMEKAHGWDYVRSKHNGRNSKK  539
N.fischeri          477 WPTEKERDRHINDKHSRAPALYKCKFAPCTYSSKRESNCKQHMEKAHGWDYVRSKHNGRSSKK  539
A.clavatus          469 WPTEKERDRHINDKHSKAPALYKCTFAPCSYQSKRESNCKQHMEKAHGWVYVRSKHNGRNSKK  531
A.oryzae            444 FPTEKETERHYNDKHCKNPRLFRCHQPGCTYASKRESNCKQHMEKTHGWVYERTKNNGKNKPV  506
A.flavus            444 FPTEKETERHYNDKHCKNPRLFRCHQPGCTYASKRESNCKQHMEKTHGWVYERTKNNGKNKPV  506
A.niger             442 WPTEKENERHVNDRHSTTPRMYACTFNGCAYKSKRESNCKQHMEKAHGWNYVRAKNNGRNAKR  504
A.terreus           443 FPTEKEKDRHFNDKHNPDPEPYECDLGGCNYRSKRLSNLKQHKEKKHGWQYVRTKSNGKRKDK  505
T.emersonii         466 WPTEKERDRHMNDRHSTSPALFKCKFAPCTYQSKRESNCKQHMEKAHGWVYVRSKNNGKNRKS  528
P.chrysogenum       453 WPTEKERDRHMNDRHSDKPTLYKCQFHKCPYASKRASNCKQHMEKSHGWVYVRSKNNGRNGSK  515
U.reesei            443 WPTEKERDRHVNDKHCKSPRLFNCRFPPCTYQSKRESNCKQHMEKAHGWVYVRSKNTGKGSDR  505
A.capsulatus        488 WPTEKERDRHVNDKHSKSPPLFKCHFSPCTYQSKRQSNCKQHMEKAHGWVYIRSKNNGKSGSR  550
C.immitis           467 WPTEKERDRHVNDKHSKSPPLYSCLFKPCTYQSKRESNCKQHMEKAHGWVYVRSKNTGKASAR  529
A.benhamiae         443 WPTEKERDRHMNDRHSKSPKKYKCLYPGCTYDTKRQSNCKQHMEKLHGYKYVRSKNNGRASHQ  505
A.gypseum           451 WPTEKERDRHMNDRHSKSPKKFKCLYPGCTYDTKRQSNCKQHMEKLHGYKYVRSKNNGRSSHQ  513
T.rubrum            443 WPTEKERDRHMNDRHSKSPKKYKCLYPGCTYDTKRQSNCKQHMEKLHGYKYVRSKNNGRASHQ  505
P.brasiliensis      508 WPTEKERDRHVNDKHSKSPQLYRCQFPPCTYQSKRDSNRKQHMEKAHGWVYVRSKNNGKASSR  570
P.nodorum           449 WPTEKERDRHMNDKHSAQPAQYKCLYPPCTYASKRESNCKQHMEKAHGWEYVRSKSNGRKKMP  511
P.tritici-repen     442 WPTEKERDRHMNDKHSAAPAQYKCLYPPCTYASKRESNCKQHMEKAHGWEYVRSKCNGRKKGP  504
C.globosum          442 WPTEKEMDRHNNDKHSASPPMYECLFKPCPYKSKRESNCKQHMEKAHGWVYVRTKTNGKKDSS  504
N.crassa            436 WPTEKEMDRHHNDKHSSAPPMYECLFKPCPYKSKRESNCKQHMEKAHGWTYVRTKTNGKKPST  498
P.anserina          444 WPTEKEMDRHHNDKHSAAPPMFECYYKPCPYKSKRESNCKQHMEKAHGWTYVRTKTNGKKPGS  506
S.macrospora        436 WPTEKEMDRHHNDKHSSAPPMYECLFKPCPYKSKRESNCKQHMEKAHGWTYVRTKTNGKKPST  498
M.grisea            437 WPTEKEMDRHHNDKHSAAPPMYECLYKPCPYKSKRESNCKQHMEKAHGWTYVRTKANGGKKID  499
T.reesei            442 WPTEKEMDRHINDKHSDAPAMYECLFKPCPYKSKRESNCKQHMEKAHGWTYVRTKTNGKKAPS  504
G.zeae              443 WPTEKEMDRHHNDKHSAAPAMYECSFKPCPYKSKRESNCKQHMEKAHGWTYVRTKTNGKKLPS  505
N.haematococca      443 WPTEKEMDRHHNDKHSDAPAMYECLFKPCPYKSKRESNCKQHMEKAHGWHYVRTKTNGKKLPG  505
V.albo-atrum        448 WPTEKEMDRHHNDKHSSAPPMHECLFQALPLQVEARVNCKQHMEKAHGWQYVRTKTNAAEAPA  510
B.cinerea           465 WPTEKEMDRHQNDKHSAAPPLFECHFKPCPYRSKRESNCKQHMEKAHGWEYIRSKNNGKNRPA  527
cons                568 :.**** :** **:*   *  . *         :   * *** ** **: * *:* ..       630

A.nidulans          517 ---PQ--Q--QTTSPSSSSVQPKQAP----S---V-WSMTPPSEAPDYR---QEPN-------  554
A.fumigatus         540 ---A-----SNGATPQTPS---IATPS-SKA---Q-GITTPLTGS--EPSPF--EPVTAY---  579
N.fischeri          540 ---A-----SNGATPQTPS---IATPS-SKA---Q-DIATPLTGS--EPSPF--EPVTAY---  579
A.clavatus          532 ---A-----STRAASQTPS---IATPS-SKA---Q-EILTPATRP--TPSPF--ESVASY---  571
A.oryzae            507 ---KQ-GS--AQPTPQSSG---IPSPA-ASHP-TG-DFSTPTTGP--TVSPS--EPPIAF---  550
A.flavus            507 ---KQ-GS--AQPTPQSSG---IPSPA-ASHP-TG-DFSTPTTGP--TVSPS--EPPIAF---  550
A.niger             505 ---RAT-S--ARASPAEESTA-HSSPL-HMS---P-DQVSPVQLAP-AQLASLQMA-------  547
A.terreus           506 ---VKGKKASPQSTPDTPG---LTTPG-TSTA--Q-SFSTPNTGP--SPSPP--QAVSRP---  551
T.emersonii         529 ---ASVS--SAQPTPQTPA---MSTPS-NS----T-DVPTPVSGP--APSPY--DPFTGI---  570
P.chrysogenum       516 ---R-GS--SSQATPQTPS---VSTPA-SKT---T-DFASPIPGP--SPSPS--DQNYNW---  557
U.reesei            506 ---G-----SPH-TPPSF------NPSPTIA---P-GIPTPISTV--ACSPY--MSPYQPQLE  545
A.capsulatus        551 ---VSGST-SGQPTPRTPN---IQTPN-SGV---M-DLPTPQSHP--GRSPY--APNVS----  593
C.immitis           530 ---G-----SNRPTPNIA------TS-PNLA---G-EMPTPISAV--APSPY--TSPYQPDFQ  569
A.benhamiae         506 ---L-----SPQPTPTTTN---ISTPQ-APS---V-DIPTPISAQ--AMSPY--LSPSEKQST  548
A.gypseum           514 ---L-----SPQPTPATTN---MNSPQ-APS---V-DIPTPISAQ--AMSPY--MSPYEQQSI  556
T.rubrum            506 ---L-----SPQPTPTTTN---ISSPQ-APS---V-DIPTPVSAQ--AMSPY--LSPSEKQST  548
P.brasiliensis      571 ---VSSS--HTPHTPQTPI---MKTPN-SGT---V-DIPTPLSHP--ALSPY--LANVS----  612
P.nodorum           512 ---TSSER--SPTTPLTPF---LGTPM-SAV------MTTPVTPY--APSPQ--VPMMD-N-F  553
P.tritici-repen     505 ---TSSDR--SPATPLTPF---IGTPM-SAN------MTTPVTPY--IPSPQ--VPMMD----  544
C.globosum          505 ---VNGNS--AHPTPQLQN---IPTPS-SEHS--M-GVATPPDTW--GPVYP--GS-------  544
N.crassa            499 LPSLGPDS--GHPTPQLQN---IGTPS-SDRS--M-SIATPSDDW--NAGLY--QT-------  541
P.anserina          507 S--IAGGS--THPTPQLGH---ISTPS-SDMS--A-GVATPPDDW--SH-IY--SS-------  546
S.macrospora        499 LPSLGPDS--GHPTPQLQN---IGTPS-SDRS--M-SIATPSDDW--NAGLY--QT-------  541
M.grisea            500 SN-PSGSV--THATPQLTN---MPTPP-SDNGAFA-GLATPPMEY--TMPAY--NN-------  543
T.reesei            505 ---QNGST--AQQTPPLAN---VSTPS-STPS--Y-SVPTPPQDQ-----V----M-------  539
G.zeae              506 ---IAGSV--QQQTPPLGN---MSTPS-SIEY--N-SVPTPPQND--VTQF----V-------  543
N.haematococca      506 ---KPASS--TQQTPPLVN---VSTPS-TTPT--YSGVPTPPQHD--VTQF----V-------  544
V.albo-atrum        511 ---LLESS--GTADPSARQ---HGN-------------------------------------  527
B.cinerea           528 AP-AVIPN--GLPTPQNTI---IHTPG-SDSN-----LESPLIAD--DEMLY--ENTSL---Y  571
cons                631              .                                                693
```

```
A.nidulans          555 ---------------------------------------------G-------------W  556
A.fumigatus         580 PPNPPFSFADP------------------------------PTQTGSGDF--PLFTT---NS  606
N.fischeri          580 PPNPPFSFADP------------------------------PTQTGSGDF--PLFTT---NS  606
A.clavatus          572 PQNPPFSFADP------------------------------PTQTASEDF--PLFTE---TG  598
A.oryzae            551 PETIPFSFADP------------------------------PVPTQTEDF--QLFSN---SP  577
A.flavus            551 PETIPFSFADP------------------------------PVPTQTEDF--QLFSN---SP  577
A.niger             548 ---------------------------------------------SDQL-----------S  552
A.terreus           552 LPTTDFNFADP------------------------------PLPTPVADF--QLFNA---NS  578
T.emersonii         571 SANQPFSFAEP------------------------------PVN-PNTDF--QLFPD---VQ  596
P.chrysogenum       558 PENPQFNFADP------------------------------PPLPHGEDF--PLFGE---TS  584
U.reesei            546 HQLL--------------------------------------------------AIASD---DT  556
A.capsulatus        594 PYDQSLQYTGPS------------GGYPLD---------DMMDPHNHDF--QLFPD---ST  628
C.immitis           570 SGTLGHN--------------------PAN----------EYNNTFNEDF--PLFPA---TP  596
A.benhamiae         549 PFEKTASDRSPSISVQDSVGYTPSPGVYPAPSPSIYSPSEVGQQNTNHDDF--MLFPD---ST  606
A.gypseum           557 PFEKPASHGSPSMSTRSSVGYAPSPGVYPGPSPSICSPSEMGQQNTTNDDF--MLFPE---ST  614
T.rubrum            549 PFEKAASDRSPSISVQDSVGYTPSPGVYPAPSPSICSPSEVGQQNTNHDDF--MLFPD---ST  606
P.brasiliensis      613 PYEQSLQYNAPGS------------GTYPLD---------DLMDPQNQDF--QLFPD---AM  648
P.nodorum           554 AEF--Y--------------------------------------GFNT--PAVSMHGFQD  571
P.tritici-repen     545 -----Y--------------------------------------GFGT--PAMSVYDYQD  559
C.globosum          545 ---------------------------------------------SLDF--PAYMP--NDA  556
N.crassa            542 ---------------------------------------------NIEF--PAYAP---EF  552
P.anserina          547 ---------------------------------------------GLEF--PTYMP---DS  557
S.macrospora        542 ---------------------------------------------SIEF--PSYAP---EF  552
M.grisea            544 ---------------------------------------------NIEF--PTYVA---QD  554
T.reesei            540 ---------------------------------------------STDF--PMYPA---DD  550
G.zeae              544 ---------------------------------------------GNDF--PLYPT---DS  554
N.haematococca      545 ---------------------------------------------GTDF--PLYPN---ES  555
V.albo-atrum        528 ------------------------------------------------------------  527
B.cinerea           572 TTN--H--------------------------------------SMDFFGPAYPA---EM  588
cons                694                                                              756

A.nidulans          557 DL--------------A--PS-----------------P--ET-P-----DL----------  567
A.fumigatus         607 PFED-LA----AGVND-FSPLPTTSLDFQAFQSQLEGADP--NGLIP-----LTFD-RQSF-D  654
N.fischeri          607 PFED-LA----AGVND-FSPLPTTSLDFQAFQSQLEGADP--NGLIP-----LAFD-RQSF-D  654
A.clavatus          599 PYAD-LA----GGVNA-FSPLPTTSLDFQAFQSQLEAADP--NGLIL----PSDFQ-RQSL-D  647
A.oryzae            578 TSL-----------G--GSPYH-NMSDAQGFP-------P--G---A----NFDLN-QPQV-P  608
A.flavus            578 TSL-----------G--GSPYH-NMSDAQGFP-------P--G---A----NFDLN-QPQV-P  608
A.niger             553 TL--------------Q--MT-----------------P--DQ-V-----SLLQM---NA-D  569
A.terreus           579 PM----------------------------------------------------------  580
T.emersonii         597 LLDN-TFDDVPMSMDD-FNPVP-TSLDFSAFQAQLEAGNP--DDLIP----AMDMY-RPSI-S  648
P.chrysogenum       585 PYL----------MTD-VNSFP-TSVNLNGFQSQFEAGDP--NGLIP----ALEMH-RQSM-N  627
U.reesei            557 ---------------------------------------P--NNN-----------------  560
A.capsulatus        629 TNNN-LFDD----FDI-NNPFA-PQLDFSAFQASLEAGDP--NEYVP----SLDMH-IPSV-P  676
C.immitis           597 NN-G---------LGENASPFT-TPFDFSSFHAGLRASDP--HEYVP----NLDIN-MPSA-T  640
A.benhamiae         607 SL------------NE-YNT---MHEDFNAFYADLQAADP--SDTLP----QLDTN-LPSA-S  645
A.gypseum           615 SL------------NE-YNS---IHDDFNAFYADLQAGDP--CDTLP----QLDMN-LPSV-S  653
T.rubrum            607 SL------------NE-YNT---MHEDFNAFYADLQAADP--SDTLP----QLDTN-LPSA-S  645
P.brasiliensis      649 DD---------------FNAFV-SPMDFSAFEAHLAATDP--NQYIP----NLDIP-MSSSAS  688
P.nodorum           572 DF-R-R----------D--SI-------------TTDGSH--FT-YSSGHSPIEPT---SF-D  600
P.tritici-repen     560 EF-R-R----------D--SV-------------TTDGSH--FT-YSSEQSPLEPT---SF-E  588
C.globosum          557 DF--------------G--LI------------------PHPQE-L-----SLDYS---PV-D  575
N.crassa            553 NF--------------N--TI------------------P--QQ-L-----ELDYS---PI-D  569
P.anserina          558 DF--------------G--MI------------------P--QE-L-----HLEYS---PV-D  574
S.macrospora        553 NF--------------N--TI------------------P--QQ-L-----ELDYS---PI-D  569
M.grisea            555 QF--------------Q-IQF------------------P--QE-L-----SLDYSPSAPS-D  575
T.reesei            551 DWLATY----------G--AQ------------------P--NT-ID----AMDL----GL-E  571
G.zeae              555 DWMS-V----------N--NI------------------P--AE-AM----NIDL----TL-D  574
N.haematococca      556 DWMS-I----------N--NI------------------P--AE-SL----DLDL----TL-D  575
V.albo-atrum        528 ------------------------------------------------------------  527
B.cinerea           589 NL--------------F--QP------------------P--QS-L-----NMEYSP-ITD-N  607
cons                757                                                              819

A.nidulans          568 ----------------------------FNTYQAPMTA-MPGSVTG-----------------  584
A.fumigatus         655 S-----GSPVP-DLIN--ETMGF----DT--------------------SPVASTDSSSL---  682
N.fischeri          655 S-----ASPVP-DLIN--GTLGF----DT--------------------SPVASTDSSSL---  682
A.clavatus          648 S-----ASPVP-DLVA--TSMGF----DG--------------------SPIASTDTSSL---  675
A.oryzae            609 GI----GSP----------------------------------------------------  613
A.flavus            609 GI----GSP----------------------------------------------------  613
A.niger             570 H-----MSPLQ-L-----S--PD----HMSPLQMTPDRASPAEMLSLQMSP-------D----  604
A.terreus           581 ----------G-GNGGVNMG-YA----DL--------------------GFPEMT--RSMDGD  605
T.emersonii         649 SNG---S-HE---SMG--SSSIF----DN--------------------SPLVMSE-NVD---  674
P.chrysogenum       628 SMSIPSAESVP-DLMG---PVSF----DG--------------------SPLTGT--ESI---  657
U.reesei            561 ------------------------------------------------------------  560
A.capsulatus        677 SSA---TTPDGTGPMG--PGTLT----DE--------------------SPFEPTH-PTP---  706
C.immitis           641 SG----ATPIN-EGFG--SA-ML----DQ--------------------SPMDLSD-SQL---  667
A.benhamiae         646 ------STPAG--NFA--AG-FQ----AQ--------------------SPLDLVG-NNA---  669
A.gypseum           654 ------STPAG--NLA--TG-FH----AQ--------------------SPPDLIG-SNA---  677
T.rubrum            646 ------STPAG--NFA--AG-FQ----AQ--------------------SPLDLVG-SNA---  669
P.brasiliensis      689 STS---TAPAS-ALLG--SG-LI----DE--------------------NPFDPRT-TTP---  716
P.nodorum           601 D-----AVTPE-D-----TA-IN----HN--------------------DV-------FN---  617
P.tritici-repen     589 D-----AVTPD-D-----TA-IN----HD--------------------DV-------YN---  605
C.globosum          576 N-----PTPST-D-----SG-MD----HT--------------------SA-------YQ---  592
N.crassa            570 N-----GTPSP-D-----SG-MD----HN--------------------SA-------YQ---  586
P.anserina          575 N-----PTPST-D-----SG-MD----HS--------------------SA-------YQ---  591
S.macrospora        570 N-----GTPLT-D-----AG-MD----HN--------------------SA-------YQ---  586
M.grisea            576 H-----ATPSS-H-----SN-VSGGSPFQ--------------------SP-------YQ---  596
T.reesei            572 N-----LSPAS-A--A--SS-YE----QY--------------------PP-------YQ---  589
G.zeae              575 S-----TSPAS-A-----SS-YE----QY--------------------AP-------YQ---  591
N.haematococca      576 S-----TSPAS-A-----SP-YE----QY--------------------AP-------YQ---  592
V.albo-atrum        528 ------------------------------------------------------------  527
B.cinerea           608 S-----FSPDG-Q-----SP-FD----AN--------------------SP-------FN---  624
cons                820                                                              882
```

```
A.nidulans          585 ----------------------TL-------------------DAVTPTTGTINSPSE-----  601
A.fumigatus         683 NFD-----------------LAWSQLDAQN--VEEEFTTLTMQMLT--------PEH-----  712
N.fischeri          683 NFD-----------------LAWSQLDAQN--VEEEFTTLTMQMLT--------PEH-----  712
A.clavatus          676 NFD-----------------LDWNQLEVQN--LDEEFTTLNMQLLT--------PPS-----  705
A.oryzae            614 ------------------------------------A-SGSSEFLT--------PPSG----  626
A.flavus            614 ------------------------------------A-SGSSEFLT--------PPSG----  626
A.niger             605 QGS-------------------SV-----QMS-PPQVS-P--AQLSTPATGPLQSPGF-----  634
A.terreus           606 V--------------------IHM-----NN--FEDMFANVNNPYLD---------------  625
T.emersonii         675 SM-------------------------------------------------------------  676
P.chrysogenum       658 NFD-----------------LDWSNLEY-P--ANEDYTAMAAQLPQ--------D-H-----  685
U.reesei            561 ----------------------------------------TMQLE----------------  565
A.capsulatus        707 NFN-----------------VDF-----DN--LDNEYTVMNMQLLT--------PAQ-----  731
C.immitis           668 DLN-----------------IDW-----HD--LENEYTAMNLQLPT--------PE------  691
A.benhamiae         670 DLD-----------------FHF-----DV--TENEYTSMNLQLLS--------PAQ-----  694
A.gypseum           678 DLD-----------------FHF-----DV--TENEYTSMNLQLLS--------PAQ-----  702
T.rubrum            670 DLD-----------------FHF-----DV--TENEYTSMNLQLLS--------PAQ-----  694
P.brasiliensis      717 NFD-----------------VDF-----DN--MDSEYITMNMQLLT--------PAQ-----  741
P.nodorum           618 ----------------------TC-----AL--GSNFT-TGFQQQPT--------PAL-----  637
P.tritici-repen     606 ----------------------CN-----LL--NFNTI-NTGFQQPT--------PMS-----  625
C.globosum          593 DIG-----------------TDF-----SM--YDDIY-NARVQLPTP--PPLLTNAY-----  622
N.crassa            587 DLN-----------------EF-----TL--IDDIY-GATVQLPNQ--VI--SPFY-----  613
P.anserina          592 DIS-----------------TDF-----TL--YEDIY-SANVQLPTP--MH--ANIY-----  619
S.macrospora        587 DLN-----------------EF-----TL--IDDIY-GATMQLPNQ--VI--SPFY-----  613
M.grisea            597 DPG-----------------SDF-----TV--YDDIY-NANAQVQMQ------NPFS-----  622
T.reesei            590 NGS-----------------TF-----IIN-DEDIY-AAHVQIPAQ--LP--TPEQVY---  619
G.zeae              592 NGS-----------------DF-----ILD-NEDLY-AAHMQLPAH--FP--SPEQAVMYS  624
N.haematococca      593 NGS-----------------AF-----ILD-NEDLY-AAHMQLPTQ--LP--TQDQMAFYN  625
V.albo-atrum        528 -------------------------------------------------------------  527
B.cinerea           625 STSVLGAQDQFQENMQAYPSNDEW-----AAYEQTDLY-SAPAQIHIP--SN--HQIY-----  672
cons                883                                                               945

A.nidulans          602 -----------PFDLAQE-------------NT-A---F------------SIQDIFPEMKA  623
A.fumigatus         713 --SVS-M---SALNSFSR-------------DP-S---I------------SNPSPLPVQK  738
N.fischeri          713 --SVP-M---SALNSFSR-------------DP-S---I------------SNPSPLPVQK  738
A.clavatus          706 --SSE-V---NALNSFSR-------------DP-S---I------------SN--ASPLQH  729
A.oryzae            627 ---ST-HSPYDPINPFPD--------------S-S---M-----------FSFEPYMQPKTE  655
A.flavus            627 ---ST-HSPYDPINPFPD--------------S-S---M-----------FSFEPYMQPKTE  655
A.niger             635 ------------VNNYSQ-------------NP-M---L-----------PLQDPYSQLKH  655
A.terreus           626 -------------------------------------------------------------  625
T.emersonii         677 ----------------E-------------VP-S---IF-DNSP--LAV---------SEN  692
P.chrysogenum       686 --SLE-A-----MKGYSN-------------DY-EQMNLNHLSYD--A--SGKASGLSPGAQ  721
U.reesei            566 ------------------------------------------------------LKDQLQ  571
A.capsulatus        732 --SVE-V---HGLRSFSR-------------NP-S---PT-CPEP--QQKTNVNHNFSPAGQ  767
C.immitis           692 --SVE-A---STLGLYSR-------------TP-S---VS-IPSP--DA-QTGVPSFSPSGQ  726
A.benhamiae         695 --SIE-A---QAMNCLPE-------------PE-N---LC-LPLP--A--LDKNSSISPAGH  728
A.gypseum           703 --SVE-A---QAMNCLSE-------------PE-N---LS-LPLP--A--LDKNSSISPAGH  736
T.rubrum            695 --SIE-A---QAMNCLPE-------------PE-N---LC-LPLP--A--LDKNSSISPAGH  728
P.brasiliensis      742 --SVE-V---HGLSSFSR-------------NT-S---VS-GLSPGSVSEQKVSSNLSPTGQ  779
P.nodorum           638 --STGFD--FDPL-PFTN-------------T-------------MSNGIPHLSPLAQ  664
P.tritici-repen     626 --T-TFD--YEPL-PFTT-------------NS-T-------------LNTGYTHLSPLAQ  653
C.globosum          623 --SKEVA---HQFMSYT-------------AA-E---LC----------------------  639
N.crassa            614 --LKDMG---QHLGAYT-------------AP-D---LC----------QPHPAHISPIGQ  642
P.anserina          620 --DKPME---PQFTPFT-------------GA-E---LC----------PLPAQLSPIGQ  647
S.macrospora        614 --PKDMG---QHLGAYT-------------AP-E---LC----------QSHPAHISPIGQ  642
M.grisea            623 --EKDAA---QFLAFVT-------------NA-E---FQ--------TTKAAPVHFSPTGQ  653
T.reesei            620 --TKMMP---QQMPVYHV-------------QQE---PC-TTV----PI-LGEPQFSPNAQ  653
G.zeae              625 -NPKMMQ---QQLPMYQQ-------------VPQQ---IP-QPI----PVQTAPSQFSPTGQ  661
N.haematococca      626 PNPKMMA---QQLPIYQA-------------PQ--------PV----TVPTGAPHFSPTGQ  658
V.albo-atrum        528 -------------------------------------------------------------  527
B.cinerea           673 --QDIMG---QNSMAFETSGVTYQTQHQQHQPQPQQ---QQ-Q------QHPQSMAHISPIGE  720
cons                946                                                              1008

A.nidulans          624 SDGLLF-PGG---------------D--MDYPDFIN-NH---NMFNDFGY---------GDF  654
A.fumigatus         739 VENSLYTPDS-C------------HVD--EGVSDLFDSG------------YQH-----KADF  769
N.fischeri          739 VENPLYTPDS-C------------HVD--EGVSDLFDPS------------YQH-----KADF  769
A.clavatus          730 SGLPFYSPSH-C------------HMD--EGVAGLQEHT------------EQK-----KLDF  760
A.oryzae            656 E-SLLF-------------------VS--GGFGDVPSM------------FEQN-----PMDF  679
A.flavus            656 E-SLLF-------------------VS--GGFGDVPSM------------FEQN-----PMDF  679
A.niger             656 EDCVLYTDND-F------------APL--NGTTDFFE-DA---S----FGYGEHMYQPGQMDF  695
A.terreus           626 ----GL-SGG-YQ------------VP--SGYE-------------------AG-----NIDF  644
T.emersonii         693 VESMGF-------------------------------------------------------  698
P.chrysogenum       722 GNAMLYSPDS-C------------NVG--DTLSERYEYGL----------QAQA-----ANDF  754
U.reesei            572 PMLTTFNPGA-FGL---------GDAHDFFGLGDAHDL------------SNTF-----GMDF  607
A.capsulatus        768 GNLMLYSPNS-QERERDLDLDLDIDMD--EGFHDGFDNYHNHNS---HIGMGKP-----NGDF  819
C.immitis           727 GNIMLYSPDS-R------------FVD--EGFADVYEP------------LKKP-----ARDF  757
A.benhamiae         729 GNVMLYSPAS-Y------------ESD--EGFDDGFDCMKM--------DIGKQ-----TGDF  763
A.gypseum           737 GNVMLYSPAS-Y------------ESD--EGFDDGFDYMKM--------DMGKQ-----TSDF  771
T.rubrum            729 GNVMLYSPAS-Y------------ESD--EGFDDGFDCMKM--------DIGKQ-----TGDF  763
P.brasiliensis      780 GNLMLYSPNS-QQ-----------NID--EGFHDAYGNHSN-------NHMGKP-----NGDF  816
P.nodorum           665 PDVTLFSPHL--------------QMD--EGFGDMDM-DM---D---MQAFSRP-----AADF  699
P.tritici-repen     654 PDVTLYSPH----------------MD--EGFGDMDM-SM--------QGFNRP-----MGDF  684
C.globosum          640 ------QPSA-------------------------------------------------CA  645
N.crassa            643 GNTMLFTPPTSLG-----------EVD--EGFEDHDF-AM--------SNCNNV-----PGDF  678
P.anserina          648 ANAMLFTPTS--------------MVD--EGFDDQHELAA--------MTNMAS-----GGDF  681
S.macrospora        643 GNTMLFTPPTSLG-----------EID--EGFDDQDF-SM--------GNCNNV-----PGDF  678
M.grisea            654 GNTMLFTPST-LA-----------DYD--ESFDDFQN-------------NAQM-----GADF  684
T.reesei            654 QNAVLYTPTS-LR-----------EVD--EGFDESY--------------AAD-----GADF  682
G.zeae              662 ETAMLFTPNS-LR-----------DVD--EGFDDSF--------------GAD-----GMDF  690
N.haematococca      659 ETAMLFTPNS-LR-----------DVD--EGFDDSF--------------GGE-----GMDF  687
V.albo-atrum        528 ---ALEQPQH--C-----------YAA--RGE------------------YQP-----LPDF  548
B.cinerea           721 GNTMLYTPKS----------------------------------------------------  730
cons               1009                                                              1071
```

```
A.nidulans          655 TMPTQGLQY-----------------GE----------TQQP-Q--FED-DSAGFLL-----D  681
A.fumigatus         770 MLFDQ--HT-NFGASSVNMSST-CQ---LSA--LHNSNQMFP-S--LNTP-D-----------  808
N.fischeri          770 MLFDQ--NA-NFGAGSVDMSST-CH---LSA--LHNANQMFP-S--LNTQ-D-----------  808
A.clavatus          761 PLYGH--QA-NLGLGAMDMSHP-CH---LPG--MQNLDQMFP-P--LDVQ-D----------S  800
A.oryzae            680 LTDPM--YNSNFM-GYE--------------------TVQP-QL--EAN-P----------N  704
A.flavus            680 LTDPM--YNSNFM-GYE--------------------TVQP-QL--EAN-P----------N  704
A.niger             696 SSIGA--FY-DDS---I--------MNS----------MLFD-S--TATGNNNNFPD-----E  726
A.terreus           645 SDL---------GY----------------------GAAVP-P--SAVF-H----------E  661
T.emersonii         699 --F-----------------------------G-----NSP----ITTS-E----------N  709
P.chrysogenum       755 TLYNQ--SA-QMGRGAHPMMQTPSD---HSAQYHQPQQQMFP-S--LEHE-R----------  796
U.reesei            608 DLYPE--IP-GNGLVD----------------------NA----------------------  622
A.capsulatus        820 TLFES--PS-TTSIDGISKH---ANFDHSLG--SQQNVHNFP-A--LDSF-ANGDGGHF---A  867
C.immitis           758 TLFDP--PL-GG-VAN---------------------QL----------------------  771
A.benhamiae         764 TLFSG--FE-NNNMCS-GLG---GN---ESE--SWNISSMFP-P--LGTL-N----------  799
A.gypseum           772 TLFSG--SE-GNNMCSTGLD---SN---DSE--GWSISSMFP-P--LGTL-N----------  808
T.rubrum            764 TLFSG--FE-NNNMCS-GLG---GN---ESE--SWNISSMFP-P--LGTL-N----------  799
P.brasiliensis      817 TLFES--PA-DTTTS---------------------AQNFP-T--LNNF-S----------D  840
P.nodorum           700 TLFDT--MQ-PSSMAMN--------NTA----------NFFP-E--FNQL-GGQFDSLYAEPS  737
P.tritici-repen     685 TLFDT--SP-PSQLPLN--------GTA----------NFFP-D--FNQL-GGQFEQMY-EPS  721
C.globosum          646 ALADR---------------------------------------------------------  650
N.crassa            679 ILYPP--TT-DAYSKPT--------FTE----------SLFA-NVDIPSM-AAGYSQPS-SQD  717
P.anserina          682 ILFPN--QA-GVS-KPM--------YND----------SLFATDLPLQGM-GTGYSQPS-TQD  720
S.macrospora        679 ILYPP--TT-GAYSKPT--------FTE----------SLFA-NVDIPSL-AAGYSQPS-SQD  717
M.grisea            685 CLFPA--NG-VKQ---N--------SPA----------PLFG-E--IPSV-AAGYSQPT-SQE  718
T.reesei            683 QLFPA--TV-DKT---D--------VFQ----------SLFT-D--MPSA-NLGFSQTT-QPD  716
G.zeae              691 PLFPG--GN-GMAKT-N--------NYQ----------PLFG-E--IPSA-NVGFSQNS-QDP  726
N.haematococca      688 PLFPS--GN-GMTNKANE-------YQQ----------PLFG-E--IPSA-NLGFSQTS-QQD  725
V.albo-atrum        549 QP-------------------------------------R-R--LPSL-------------  556
B.cinerea           731 ---------------------------------------------------------------  730
cons               1072                                                               1134

A.nidulans          682 ----V--YN-DMH----------TYG-INPGPGGL  698
A.fumigatus         809 LNY-M--TQ-GWP------------Q-DMEM-EHF  825
N.fischeri          809 LNY-M--TQ-GWP------------Q-DMEM-EHF  825
A.clavatus          801 IAY-M--NQ-GWP------------Q-DVDM-DLF  817
A.oryzae            705 AQF-G--LD-GWD----------DFM-CKPDQSSM  724
A.flavus            705 AQF-G--LD-GWD----------DFM-CKPDQSSM  724
A.niger             727 ----------SMY----------M-F-NQTQMPPF  739
A.terreus           662 DQF-LPETI-NWE-------EFAMLN--GMDTMSK  685
T.emersonii         710 IKF-D---S-EWAH------NWYQSG-D-QYPYPL  731
P.chrysogenum       797 AMQ-S--MQ-PWNGQQPQGHYL--PR-DMDLEFMK  824
U.reesei            623 --------S-PFAQ------DF------DWQTFLC  636
A.capsulatus        868 DSD-A--AR-GWPADQIDPMDLHRDV-DEYIMGGF  897
C.immitis           772 ----------GMPL------DCSENA-FFTWYNES  789
A.benhamiae         800 DQL-N--NA-NWPQ-QLGSREIDIDS-IMEMDDTF  828
A.gypseum           809 DQL-N--NA-DWPQ-QLGSREMDIDS-IME-IDDI  836
T.rubrum            800 DQL-N--NA-NWPQ-QLGSREIDIDS-IMEMDGTF  828
P.brasiliensis      841 DHY-S-ADD-AWGV-QIDPMDMHREA-D-EYMMDY  869
P.nodorum           738 ----T--TLDDL----------------MGNNYGA  750
P.tritici-repen     722 ----T--TLDDLN----------Y----MYFSNQQ  736
C.globosum          651 -------PG-EYH----------A----LHSQLSC  663
N.crassa            718 ILHAY--QG-DWT----------S----HDMNAYF  735
P.anserina          721 LINGF--HV-DWS----------AHD--LSAYLPQ  740
S.macrospora        718 ILHAY--QG-DWT----------S----HDMNGYF  735
M.grisea            719 LL-----PN-DWR----------MNQYQGLGFSQN  737
T.reesei            717 IF--N--QI-DWS----------NLD---YQGFQE  733
G.zeae              727 -F--Q--MM-DWS----------SG----GFLANL  741
N.haematococca      726 -L--F--QM-DLS----------NMD--FQQNFTD  742
V.albo-atrum        557 ----------------------R----PGRGVC  563
B.cinerea           731 ------------------------------MQDF  734
cons               1135                                   1169
```

Appendix 8 – Control genes from *A. nidulans* used to analyse the frequency of random occurrences of TF binding site motifs within promoters.

Genes were chosen at random from the *A. nidulans* genome by selecting every 125th gene based on accession numbers at the Broad Institute website (www.broadinstitute.org/). Some genes were not available for analysis or had incomplete promoter sequences available so were excluded from the analysis. The control group comprised 66 genes. All 66 of the corresponding promoters (2 kb) were used to analyse the frequency of random occurrences of the StzA/Ace1 binding site motif (5'-AGGCA). A smaller group of control promoters was obtained (N = 22) by selecting every third gene available (marked with an asterisk), and used to analyse the frequency of random occurrences of experimentally characterised binding site motifs for a variety of TFs.

Accession number	Known or predicted function	No. of potential StzA binding sites
AN0125 (not present)	-	-
AN0250	MFS monosaccharide transporter, putative	7
AN0375 (not present)	-	-
AN0500	Conserved hypothetical protein	5
AN0625*	NOL1/NOP2/sun domain-containing protein	1
AN0750	Hypothetical protein	2
AN0875	Homoserine acetyltransferase	3
AN1000*	Conserved hypothetical protein	5
AN1125 (not present)	-	-
AN1250	Conserved hypothetical protein	9
AN1375	Conserved hypothetical protein	2
AN1500*	C_2H_2 finger domain protein, putative	3
AN1625	Conserved hypothetical protein	3
AN1750	ATP-dependent RNA helicase Mak5, putative	3
AN1875*	C_2H_2 finger domain-containing protein	2
AN2000	Polyubiquitin	3
AN2125 (not present)	-	-
AN2250	Conserved hypothetical protein	5
AN2375*	Conserved hypothetical protein	4
AN2500	Nicotinamide N-methyltransferase Nnt1, putative	3
AN2625	Conserved hypothetical protein	6
AN2750*	Conserved hypothetical protein	9
AN2875	Fructose-bisphosphate aldolase	5
AN3000	Conserved hypothetical protein	3
AN3125*	RTA1 domain-containing protein	7
AN3250	Putative $Zn(II)_2Cys_6$ transcription factor	3
AN3375	DnaJ domain-containing protein	5
AN3500 (not present)	-	-
AN3625*	Conserved hypothetical protein	1
AN3750	Conserved hypothetical protein	7
AN3875	Conserved hypothetical protein	4
AN4000*	Polyadenylate-binding protein	6
AN4125	Conserved hypothetical protein	6
AN4250	mitochondrial carrier protein (Ymc1), putative	8
AN4375*	Conserved hypothetical protein	6
AN4500	Conserved hypothetical protein	9
AN4625	Arsenic methyltransferase Cyt19	6
AN4750*	Conserved hypothetical protein	6
AN4875	YagE family protein	7
AN5000	Hypothetical membrane protein, putative	5
AN5125*	Hypothetical protein	5
AN5250 (not present)	-	-
AN5375	Conserved hypothetical protein	4
AN5500	Saponin hydrolase	3
AN5625*	Conserved hypothetical protein	3
AN5750	Conserved hypothetical protein	4
AN5875	Ribosomal RNA assembly protein mis3	8
AN6000*	Polyketide synthase, putative	4
AN6125	Nuclear distribution protein NudE	5
AN6250	Arginine-tRNA-protein transferase 1	0
AN6375 (not present)	-	-
AN6500*	60S ribosomal protein L28	6
AN6625	Conserved hypothetical protein	5
AN6750	Conserved hypothetical protein	1
AN6875*	Kinesin family protein	3
AN7000	Succinyl-CoA synthetase beta subunit, putative	2
AN7125 (not present)	-	-
AN7250	Na^+/H^+ exchanger AnNHA1	3
AN7375*	Cyclopropane-fatty-acyl-phospholipid synthase	9
AN7500	64 kDa mitochondrial NADH dehydrogenase	5
AN7625	Myo-inositol-1-phosphate synthase	4
AN7750*	DUF625 domain protein, putative	1
AN7875	Conserved hypothetical protein	2
AN8000 (not analysed) Only 28 bp of promoter sequence available.	Histone acetylase complex subunit Paf400	-
AN8125	Fungal specific transcription factor domain-containing protein	3
AN8250*	Cytochrome P450, putative	3
AN8375	Conserved hypothetical protein	5
AN8500	Conserved hypothetical protein	6
AN8625*	Conserved hypothetical protein	2
AN8750	Conserved hypothetical protein	2
AN8875	UPD-GlcNAc transporter	2
AN9000*	MFS transporter, putative	2
AN9125	DNA-directed RNA polymerase III subunit Rpc34	9
AN9250	O-acetyltransferase, putative	2
AN9375*	Taurine catabolism dioxygenase TauD	3
AN9500 (not analysed) Only 1087 bp of promoter available.	Conserved hypothetical protein	-

Appendix 9 – Numbers of binding motifs for a variety of TFs within the promoters of (a) *stzA* genes, and (b) a control group of genes. Nineteen *stzA* promoters were used in the analysis. The control group comprised 22 genes. The Microsoft Word® "Find" tool was used to identify all occurrences of TF binding sequence permutations. Reverse complements of sequences were deduced to enable identification of binding site motifs on complementary DNA strands. The number of TF motifs required in each 2 kb promoter to demonstrate significant enrichment is shown in brackets following the name of the TF. Promoters that are significantly enriched for a particular TF binding site motif are represented by grey highlights.

a) *stzA* promoters

Transcription factor	Species																		
	A. nidulans	***A. fumigatus***	***N. fischeri***	***A. clavatus***	***A. oryzae***	***A. flavus***	***A. niger***	***A. capsulatus***	***A. terreus***	***C. immitis***	***U. reesii***	***P. nodorum***	***P. tritici-repentis***	***T. reesei***	***M. grisea***	***C. globosum***	***N. crassa***	***P. anserina***	***P. chrysogenum***
AbaA (5)	7	3	4	4	5	5	1	3	1	2	3	1	1	0	3	2	3	3	4
AflR (3)	0	0	0	0	0	0	1	0	2	0	0	0	0	0	1	1	0	0	0
AlcR (14)	6	3	3	12	6	6	10	16	7	7	7	17	9	14	9	10	7	10	8
AmdA/X (5)	1	0	0	2	3	3	1	4	2	1	0	5	2	1	5	4	3	4	4
AmyR (3)	3	2	0	2	0	0	1	3	1	1	0	2	4	2	1	2	0	1	1
AnCF (8)	3	2	3	2	11	11	8	5	6	9	4	3	9	3	10	3	7	8	10
AreA (10)	4	5	6	10	9	9	7	6	6	2	6	6	6	2	5	6	4	8	7
AtfA (2)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BrlA (5)	2	3	5	5	3	3	2	2	1	5	0	4	2	2	3	3	3	6	2
CDRE (5)	4	2	2	2	3	3	5	2	3	4	1	3	0	2	1	2	1	1	4
CpcA (2)	1	1	1	1	2	2	1	1	1	1	1	1	1	2	1	1	1	1	1
CreA (14)	10	2	3	5	6	6	14	14	11	9	5	9	10	15	9	14	11	8	12
FarA/FarB (3)	0	1	1	0	1	1	1	0	0	1	0	1	0	1	1	0	1	0	0
HacA (1)	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
MeaB (1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NirA (1)	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
PacC (5)	1	3	2	0	3	3	2	6	1	3	2	2	3	3	5	3	4	4	1
PecR (3)	3	1	2	2	0	0	1	0	1	2	0	0	1	0	0	0	0	2	1
RlmA (3)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
STRE 8)	2	4	3	3	4	4	7	5	1	5	1	4	3	2	2	4	3	9	2
StuA (5)	0	0	0	0	0	0	0	0	3	0	2	3	2	1	2	5	1	2	0
StzA (8)	4	7	4	4	4	4	3	3	5	3	3	5	2	2	3	2	5	3	6
XlnR (5)	1	2	1	1	2	2	2	0	1	0	1	3	2	0	1	1	2	0	0

b) Control group promoters

	Accession number																					
Transcription factor	**AN0625**	**AN1000**	**AN1500**	**AN1875**	**AN2375**	**AN2750**	**AN3125**	**AN3625**	**AN4000**	**AN4375**	**AN4750**	**AN5125**	**AN5625**	**AN6000**	**AN6500**	**AN6875**	**AN7375**	**AN7750**	**AN8250**	**AN8625**	**AN9000**	**AN9375**
AbaA (5)	2	3	0	0	4	2	1	0	4	2	6	4	1	3	2	2	3	3	7	4	3	2
AflR (3)	1	1	2	0	2	0	1	1	0	1	0	0	1	0	0	0	0	0	1	1	0	2
AlcR (14)	12	13	9	8	8	2	2	10	9	7	5	8	2	2	10	9	11	9	6	11	8	9
AmdA/X (5)	1	3	3	5	3	1	2	1	2	2	2	2	1	2	1	2	4	2	0	2	4	2
AmyR (3)	2	0	2	0	1	0	3	0	2	0	2	1	1	2	0	0	1	0	0	1	1	2
AnCF (8)	2	2	8	4	8	6	4	6	5	5	4	5	4	4	8	4	9	2	5	7	5	5
AreA (10)	3	8	3	7	4	8	1	8	7	4	3	4	1	4	3	4	2	3	6	5	4	4
AtfA (2)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BrlA (5)	3	1	2	4	3	1	4	1	1	3	0	5	4	2	1	1	1	3	1	1	2	1
CDRE (5)	4	0	1	3	2	1	2	1	1	2	1	1	0	1	5	1	3	0	3	1	4	1
CpcA (2)	0	1	0	0	1	0	1	1	0	1	0	1	0	0	1	0	0	0	0	0	0	0
CreA (14)	13	9	7	10	11	5	10	14	7	4	7	4	6	9	10	8	8	15	5	10	14	8
FarA/FarB (3)	1	0	1	0	0	0	2	2	1	0	1	2	1	0	0	0	0	4	1	2	3	0
HacA (1)	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
MeaB (1)	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
NirA (1)	0	1	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0
PacC (5)	0	1	7	2	5	3	2	6	3	3	3	1	1	2	4	1	4	4	4	0	1	7
PecR (3)	0	0	0	3	2	2	1	0	0	2	2	3	0	0	2	1	1	2	1	1	2	0
Rlm1 (3)	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	2	0	0	0
STRE (8)	2	4	8	6	3	1	4	1	3	6	0	5	5	3	2	3	1	4	4	4	2	4
StuA (5)	2	0	1	2	3	2	0	2	2	0	4	0	4	4	3	3	1	1	0	4	1	1
StzA (8)	1	5	3	2	4	9	7	1	6	6	6	5	3	4	6	3	9	1	3	2	2	3
XlnR (5)	1	0	1	0	3	1	3	2	1	0	0	0	0	1	3	2	1	1	1	0	1	2

Appendix 10 – Chi-square analysis of StzA binding site motif frequencies in promoters of functionally related groups of genes. All 66 promoters (2 kb) from the control group of genes were used to determine expected frequencies of StzA binding site motifs.

Osmotic stress group (χ^2 = 42.98)

No. of StzA binding site motifs	Control group frequencies	Osmotic stress group		
		Observed (O)	Expected (E)	$\frac{(O-E)^2}{E}$
0, 1 or 2	15	10	13.35	0.841
3	15	8	13.35	2.144
4	6	19	5.34	34.943
5	11	8	9.79	0.327
6	8	6	7.12	0.176
7	4	6	3.56	1.672
8, 9 or 10	7	2	6.23	2.872
Total	**66**	**59**	**58.74**	**42.975 (χ^2)**

Note that the expected value 3.56 is less than 5 but accounts for less than one-fifth of categories, so meets the criteria of the test. The critical value is 12.59 for 6 df at the 5% significance level.

DNA repair group (χ^2 = 15.74)

No. of StzA binding site motifs	Control group frequencies	DNA repair group		
		Observed (*O*)	Expected (*E*)	$\frac{(O-E)^2}{E}$
0 or 1	5	7	7.50	0.033
2	10	10	15.00	1.667
3	15	22	22.50	0.011
4	6	17	9.00	7.111
5	11	13	16.50	0.742
6	8	19	12.00	4.083
7	4	5	6.00	0.167
8, 9 or 10	7	6	10.50	1.929
Total	**66**	**99**	**99.00**	**15.743 (χ^2)**

The critical value is 14.07 for 7 df at the 5% significance level.

Cation homeostasis group (χ^2 = 1.92)

No. of StzA binding site motifs	Control group frequencies	Cation homeostasis group		
		Observed (*O*)	Expected (*E*)	$\frac{(O-E)^2}{E}$
0, 1 or 2	15	4	5.70	0.507
3	15	6	5.70	0.016
4 or 5	17	5	6.46	0.330
6, 7, 8, 9 or 10	19	10	7.22	1.070
Total	**66**	**25**	**25.08**	**1.923 (χ^2)**

The critical value is 7.81 for 3 df at the 5% significance level.

***REALALE*-containing group (χ^2 = 31.22)**

No. of StzA binding site motifs	Control group frequencies	*REALALE*-containing group		
		Observed (*O*)	Expected (*E*)	$\frac{(O-E)^2}{E}$
0, 1 or 2	15	3	10.20	5.082
3	15	4	10.20	3.769
4 or 5	17	9	11.56	0.567
6 or 7	12	16	8.16	7.533
8, 9 or 10	7	13	4.76	14.264
Total	**66**	**45**	**44.88**	**31.215 (χ^2)**

An expected value of less than 5 is only present for only one category out of the five, so meets the criteria of the test. The critical value is 9.49 for 4 df at the 5% significance level.

References

ABRAHAM, R. T. (2001) Cell cycle checkpoint signaling through the ATM and ATR kinases. *Genes & Development*, **15:** 2177–2196.

ALTSCHUL, S. F., MADDEN, T. L., SCHÄFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25:** 3389–3402.

ANDERSEN, M. R. & NIELSEN, J. (2009) Current status of systems biology in *Aspergilli*. *Fungal Genetics & Biology*, **46:** S180–S190.

ANDRIANOPOULOS, A. & TIMBERLAKE, W. E. (1994) The *Aspergillus nidulans abaA* gene encodes a transcriptional activator that acts as a genetic switch to control development. *Molecular & Cellular Biology*, **14:** 2503–2515.

ANDRIANOPOULOS, A., BRONS, J., DAVIS, M. A. & HYNES, M. J. (1997) The *amdA* regulatory gene of *Aspergillus nidulans*: characterization of gain-of-function mutations and identification of binding sites for the gene product. *Fungal Genetics & Biology*, **21:** 50–63.

ARCHER, D. B. & DYER, P. S. (2004) From genomics to post-genomics in *Aspergillus*. *Current Opinion in Microbiology*, **7:** 499–504.

ARNAUD, M. B., CHIBUCUS, M. C., COSTANZO, M .C., CRABTREE, J., INGLIS, D. O., LOTIA, A., ORVIS, J., SHAH, P., SKRZYPEK, M. S., BINKLEY, G., MIYASATO, S. R., WORTMAN, J. R. & SHERLOCK, G. (2010) The *Aspergillus* Genome Database, a curated comparative genomics resource for gene, protein and sequence information for the *Aspergillus* research community. *Nucleic Acids Research*, **38:** D420–427.

ARO, N., ILMÉN, M., SALOHEIMO, A. & PENTTILÄ, M. (2003) ACE1 of *Trichoderma reseei* is a repressor of cellulase and xylanase expression. *Applied and Environmental Biology,* **69:** 56–65.

ARST, H. N. & PEÑALVA, M. A. (2003) pH regulation in *Aspergillus* and parallels with higher eukaryotic regulatory systems. *Trends in Genetics*, **19:** 224−231.

ARST, H. N. & PEÑALVA, M. A. (2004) Recent advances in the characterization of ambient pH regulation of gene expression in filamentous fungi and yeasts. *Annual Review of Microbiology*, **58:** 425−451.

ATTWELL, N. A., CLEMENT, D. J., STANLEY, M. S., CLIPSON, N. J. W., HOOLEY, P. & FINCHAM, D. A. (1995) Arginine interactions with the Na^+/H^+ antiporter in protoplasts from *Aspergillus nidulans*. *Journal of Physiology*, **489:** 104−105.

BAGINSKI, M. & CZUB, J. (2009) Amphotericin B and its new derivatives – mode of action. *Current Drug Metabolism*, **10:** 459–469.

BAKER, B. J., LUTZ, M. A., DAWSON, S. C., BOND, P. L. & BANFIELD, J. F. (2004) Metabolically active eukaryotic communities in extremely acidic mine drainage. *Applied & Environmental Microbiology*, **70:** 6264–6271.

BALÁZS, A., PÓCSI, I., HAMARI, Z., LEITER, E., EMRI, T., MISKEI, M., OLÁH, J., TÓTH, V., HEGEDÚS, N., PRADE, R. A., MOLNÁR, M. & PÓCSI, I. (2010) AtfA bZIP-type transcription factor regulates oxidative and osmotic stress responses in *Aspergillus nidulans*. *Molecular Genetics & Genomics*, **283:** 289–303.

BALLANCE, D. J. & TURNER, G. (1985) Development of a high-frequency transforming vector for *Aspergillus nidulans*. *Gene*, **36:** 321–331.

BARHAM-MORRIS, J. B. (2006) *Characterisation of the C_2H_2 transcription factor StzA in Aspergillus nidulans.* Ph.D. Thesis, University of Wolverhampton.

BATADA, N. N., URRUTIA, A. O. & HURST, L. D. (2007) Chromatin remodelling is a major source of co-expression of linked genes in yeast. *Trends in Genetics*, **23:** 480–484.

BAYRAM, Ö., BIESEMANN, C., KRAPPMANN, S., GALLAND, P. & BRAUS, G. H. (2008) More than a repair enzyme: *Aspergillus nidulans* photolyase-like CryA is a regulator of sexual development. *Molecular Biology of the Cell*, **19:** 3254–3262.

BEAUREGARD, A., CURCIO, M. J. & BELFORT, M. (2008) The take and give between retrotransposable elements and their hosts. *Annual Review of Genetics*, **42:** 587–617.

BEEVER, R. E. & LARACY, E. P. (1986) Osmotic adjustment in the filamentous fungus *Aspergillus nidulans*. *Journal of Bacteriology*, **168:** 1358–1365.

BENČINA, M., BAGAR, T., LAH, L. & KRAŠEVEC, N. (2009) A comparative genomic analysis of calcium and proton signalling/homeostasis in *Aspergillus* species. *Fungal Genetics & Biology*, **46:** S93–S104.

BÉNÉDETTI, H., RATHS, S., CRAUSAZ, F. & RIEZMAN, H. (1994) The END3 gene encodes a protein that is required for the internalization step of endocytosis and for actin cytoskeleton organization in yeast. *Molecular Biology of the Cell*, **5:** 1023–1037.

BENITO, B., GARCIADEBLÁS, B. & RODRÍGUEZ-NAVARRO, A. (2002) Potassium- or sodium-efflux ATPase, a key enzyme in the evolution of fungi. *Microbiology*, **148:** 933–941.

BERBEE, M. L. (2001) The phylogeny of plant and animal pathogens in the Ascomycota. *Physiological and Molecular Plant Pathology*, **59:** 165–187.

BLOMBERG, A. & ADLER, L. (1992) Physiology of osmotolerance in fungi. *Advances in Microbial Physiology*, **33:** 145–212.

BORNEMAN, A. R., HYNES, M. J. & ANDRIANOPOULOS, A. (2000) The *abaA* homologue of *Penicillium marneffei* participates in two developmental programmes: conidiation and dimorphic growth. *Molecular Microbiology*, **38:** 1034–1047.

BORNEMAN, A. R., GIANOULIS, T. A., ZHANG, Z. D., YU, H., ROZOWSKY, J., SERINGHAUS, M. R., WANG, L. Y., GERSTEIN, M. & SNYDER, M. (2007) Divergence of transcription factor binding sites across related yeast species. *Science*, **317:** 815–819.

BORSUK, P., DZIKOWSKA, A., EMPEL, J., GRZELAK, A., GRZESKOWIAK, R. & WEGLENSKI, P (1999) Structure of the arginase coding gene and its transcript in *Aspergillus nidulans*. *Acta Biochimica Polonica*, **46:** 391–405.

BRACHAT, S., DIETRICH, F. S., VOEGELI, S., ZHANG, Z., STUART, L., LERCH, A., GATES, K., GAFFNEY, T. & PHILIPPSEN, P. (2003) Reinvestigation of the *Saccharomyces cerevisiae* genome annotation by comparison to the genome of a related fungus: *Ashbya gossypii*. *Genome Biology*, **4:** R45.

BRADFORD, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, **72:** 248–254.

BRAUS, G. H., SASSE, C. & KRAPPMANN, S. (2006) Amino acid acquisition, cross-pathway control, and virulence in *Aspergillus. Medical Mycology*, **44:** S91–S94.

BRETTEL, K. & BYRDIN, M. (2010) Reaction mechanisms of DNA photolyase. *Current Opinion in Structural Biology*, **20:** 693–701.

BROWN, R. S. (2005) Zinc finger proteins: getting a grip on RNA. *Current Opinion in Structural Biology*, **15:** 94–98.

BRUDNO, M., DO, C. B., COOPER, G. M., KIM, M. F., DAVYDOV, E., GREEN, E. D., SIDOW, A. & BATZOGLOU, S. (2003) LAGAN and Multi-LAGAN: Efficient tools for large-scale multiple alignment of genomic DNA. *Genome Research*, **13:** 721–731.

BURGER, G., TILBURN, J. & SCAZZOCCHIO, C. (1991) Molecular cloning and functional characterization of the pathway-specific regulatory gene *nirA*, which controls nitrate assimilation in *Aspergillus nidulans*. *Molecular & Cellular Biology*, **11:** 795–802.

CARUSO, M. L., LITZKA, O., MARTIC, G., LOTTSPEICH, F. & BRAKHAGE, A. A. (2002) Novel basic-region helix-loop-helix transcription factor (AnBH1) of *Aspergillus nidulans* counteracts the CCAAT-binding complex AnCF in the promoter of a penicillin biosynthesis gene. *Journal of Molecular Biology*, **323:** 425–439.

CARVALHO, N. D. S. P., ARENTSHORST, M., KWON, M. J., MEYER, V. & RAM, A. F. J. (2010) Expanding the *ku70* toolbox for filamentous fungi: establishment of complementation vectors and recipient strains for advanced gene analysis. *Applied Microbiology & Biotechnology*, **87:** 1463–1473.

CARY, J. W., KLICH, M. A. & BELTZ, S. B. (2005) Characterization of aflatoxin-producing fungi outside of Aspergillus section Flavi. *Mycologia*, **97:** 425–432.

CHANG, Y. C. & TIMBERLAKE, W. E. (1993) Identification of *Aspergillus brlA* response elements (BREs) by genetic selection in yeast. *Genetics*, **133:** 29–38.

CHAVEROCHE, M., GHIGO, J. & D'ENFERT, C. (2000) A rapid method for efficient gene replacement in *Aspergillus nidulans*. *Nucleic Acids Research*, **28:** e97.

CHEN, R. E. & THORNER, J. (2007) Function and regulation in MAPK signaling pathways: Lessons learned from the yeast *Saccharomyces cerevisiae*. *Biochimica et Biophysica Acta – Molecular Cell Research*, **1773:** 1311–1340.

CHEN, D., TOONE, W. M., MATA, J., LYNE, R., BURNS, G., KIVINEN, K., BRAZMA, A., JONES, N. & BÄHLER, J. (2003) Global transcriptional responses of fission yeast to environmental stress. *Molecular Biology of the Cell*, **14:** 214–229.

CHENNA, R., SUGAWARA, H., KOIKE, T., LOPEZ, R., GIBSON, T. J., HIGGINS, D. G. & THOMPSON, J. D. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic acids research,* **31:** 3497–3500.

CHILTON, I. J., DELANEY, C. E., BARHAM-MORRIS, J., FINCHAM, D. A., HOOLEY, P. & WHITEHEAD, M. P. (2008) The *Aspergillus nidulans* stress response transcription factor StzA is ascomycete-specific and shows species-specific polymorphisms in the C-terminal region. *Mycological Research*, **112:** 1435–1446.

CHO, J. H., YUN, S. S., JANG, Y. K., CHA, M. J., KWON, N. J. & CHAE, S-K. (2003) Identification and cloning of *jipA* encoding a polypeptide that interacts with a homolog of yeast Rad6, UVSJ, in *Aspergillus nidulans. Journal of Microbiology*, **41:** 46–51.

CHRISTENSEN, U., GRUBEN, B. S., MADRID, S., MULDER, H., NIKOLAEV, I. & DE VRIES, R. P. (2011) Unique regulatory mechanism for D-galactose utilization in *Aspergillus nidulans*. *Applied & Environmental Microbiology*, **77:** 7084–7087

CHU, G. (1997) Double strand break repair. *Journal of Biological Chemistry*, **272:** 24097–24100.

CLEMENT, D. J., STANLEY, M. S., ATTWELL, N. A., CLIPSON, N. J., FINCHAM, D. A. & HOOLEY, P. (1996) Evidence for *sltA1* as a salt-sensitive allele of the arginase gene (*agaA*) in the ascomycete *Aspergillus nidulans*. *Current Genetics*, **29:** 462–467.

CLEMENT, D. J., STANLEY, M. S., O'NEIL, J., WOODCOCK, N. A., FINCHAM, D. A., CLIPSON, N. J. W. & HOOLEY, P. (1999) Complementation cloning of salt tolerance determinants from the marine hyphomycete *Dendryphiella salina* in *Aspergillus nidulans*. *Mycological Research*, **103:** 1252–1258.

CLIPSON, N. J. W. & JENNINGS, D. H. (1992) *Dendryphiella salina* and *Debaryomyces hansenii*: models for ecophysiological adaptation to salinity by fungi that grow in the sea. *Canadian Journal of Botany*, **70:** 2097–2105.

CLIPSON, N. J. W. & JENNINGS, D. H. & SMITH, J. L. (1989) The response to salinity at the microscopic level of the marine fungus *Dendryphiella salina*. *New Phytologist*, **113:** 21–27.

CLUTTERBUCK, A. J. (1974) *Aspergillus nidulans*. In *Handbook of Genetics* (Ed, King, R. C.). Vol. 1, 447–510. Plenum Press, New York.

CLUTTERBUCK, A. J. (1997) The validity of the *Aspergillus nidulans* linkage map. *Fungal Genetics & Biology*, **21:** 267–277.

COUTINHO, P. M., ANDERSEN, M. R., KOLENOVA, K., VAN KUYK, P. A., BENOIT, I., GRUBEN, B. S., TREJO-AGUILAR, B., VISSER, H., VAN SOLINGEN, P., PAKULA, T., SEIBOTH, B., BATTAGLIA, E., AGUILAR-OSORIO, G., DE JONG, J. F., OHM, R. A., AGUILAR, M., HENRISSAT, B., NIELSEN, J., STÅLBRAND, H., & DE VRIES, R. P. (2009) Post-genomic insights into the plant polysaccharide degradation potential of *Aspergillus nidulans* and comparison to *Aspergillus niger* and *Aspergillus oryzae*. *Fungal Genetics & Biology*, **46:** S161–169.

COVE, D. J. (1976) Chlorate toxicity in *Aspergillus nidulans*: Studies of mutants altered in nitrate assimilation. *Molecular & General Genetics*, **146:** 147–159.

CUBERO, B. & SCAZZOCCHIO, C. (1994) Two different, adjacent and divergent zinc finger-binding sites are necessary for CREA-mediated carbon-catabolite repression in the proline gene cluster of *Aspergillus nidulans*. *EMBO Journal*, **13:** 407–415.

CUOMO, C. A. & BIRREN, B. W. (2010) The fungal genome initiative and lessons learned from genome sequencing. *Methods in Enzymology*, **470:** 833–855.

DABOUSSI, M-J. & CAPY, P. (2003) Transposable elements in filamentous fungi. *Annual Reviews in Microbiology*, **57:** 275–299.

DAMVELD, R. A., ARENTSHORST, M., FRANKEN, A., VAN KUYK, P. A., KLIS, F. M., VAN DEN HONDEL, C. A. M. J. J. & RAM, A. F. J. (2005) The *Aspergillus niger* MADS-box transcription factor RlmA is required for cell wall reinforcement in response to cell wall stress. *Molecular Microbiology*, **58:** 305–319.

DAVIES, J. M., BROWNLEE, C. & JENNINGS, D. H. (1990) Electrophysiological evidence for an electrogenic proton pump and the proton symport of glucose in the marine fungus *Dendryphiella salina*. *Journal of Experimental Botany*, **41:** 449–456.

DAVIS, D. J., BURLAK, C. & MONEY, N. P. (2000a) Osmotic pressure of fungal compatible osmolytes. *Mycological Research*, **104:** 800–804.

DAVIS, C. A., GRATE, L., SPINGOLA, M. & ARES, M. (2000b) Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. *Nucleic Acids Research*, **28:** 1700–1706.

DEREEPER, A., GUIGNON, V., BLANC, G., AUDIC, S., BUFFET, S., CHEVENET, F., DUFAYARD, J. -F., GUINDON, S., LEFORT, V., LESCOT, M., CHAVERIE, J. -M. & GASCUEL, O. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, **36:** W465–469.

DE SOUZA, C. P. C., YE, X. S. & OSMANI, S. A. (1999) Checkpoint defects leading to premature mitosis also cause endoreplication of DNA in *Aspergillus nidulans*. *Molecular Biology of the Cell*, **10:** 3661–3674.

DE VRIES, R. P. & VISSER, J. (2001) *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides. *Microbiology & Molecular Biology Reviews*, **65:** 497–522.

DE VRIES, R. P., VAN DE VONDERVOORT, P. J., HENDRIKS, L., VAN DE BELT, M. & VISSER, J. (2002) Regulation of the alpha-glucuronidase-encoding gene (*aguA*) from *Aspergillus niger*. *Molecular Genetics & Genomics*, **268:** 96–102

DE VRIES, R. P., FLITTER, S. J., VAN DE VONDERVOORT, P. J. I., CHAVEROCHE, M. K., FONTAINE, T., FILLINGER, S., RUIJTER, G. J. G. D'ENFERT, C. & VISSER, J. (2003) Glycerol dehydrogenase, encoded by *gldB*, is essential for osmotolerance in *Aspergillus nidulans*. *Molecular Microbiology*, **49:** 131–141.

DEZOTTI, N. O. C. R. & ZUCCHI, T. M. A. D. (2001) Identification of *Aspergillus nidulans* genes essential for the accumulation of sterigmatocystin. *Fungal Genetics & Biology*, **34:** 93–105.

DINAMARCO, T. M., PIMENTEL, C., SAVOLDI, M., MALAVAZI, I., SORIANI, F. M., UYEMURA, S. A., LUDOVICO, P., GOLDMAN, M. H. & GOLDMAN, G. H. The roles played by *Aspergillus nidulans* apoptosis-inducing factor (AIF)-like mitochondrial oxidoreductase (AifA) and NADH-ubiquinone oxidoreductases (NdeA-B and NdiA) in farnesol resistance. *Fungal Genetics & Biology*, **47:** 1055–1069.

DINGWALL, C. & LASKEY, R. A. (1998) Nuclear import: a tale of two sites. *Current Biology*, **8:** 922–924.

DONIGER, S. W. & FAY, J. C. (2007) Frequent gain and loss of functional transcription factor binding sites. *PLoS Computational Biology*, **3:** e99.

DONIGER, S. W., HUH, J. & FAY, J. C. (2005) Identification of functional transcription factor binding sites using closely related *Saccharomyces* species. *Genome Research*, **15:** 701–709.

DOWZER, C. E. A. & KELLY, J. M. (1991) Analysis of the *creA* gene, a regulator of carbon catabolite repression in *Aspergillus nidulans*. *Molecular & Cellular Biology*, **11:** 5701−5709.

DUNLAP, J. C., BORKOVICH, K. A., HENN, M. R., TURNER, G. E., SACHS, M. S., GLASS, N. L., McCLUSKEY, K., PLAMANN, M., GALAGAN, J. E., BIRREN, B. W., WEISS, R. L., TOWNSEND, J. P., LOROS, J. J., NELSON, M. A., LAMBREGHTS, R., COLOT, H. V., PARK, G., COLLOPY, P., RINGELBERG, C., CREW, C., LITVINKOVA, L., DECAPRIO, D., HOOD, H. M., CURILLA, S., SHI, M., CRAWFORD, M., KOERHSEN, M., MONTGOMERY, P., LARSON, L., PEARSON, M., KASUGA, T., TIAN, C., BAŞTÜRKMEN, M., ALTAMIRANO, L. & XU, J. (2007) Enabling a community to dissect an organism: overview of the *Neurospora* functional genomics project. *Advances in Genetics,* **57:** 49–96.

DUTTON, J. R., JOHNS, S. & MILLER, B. L. (1997) StuAp is a sequence-specific transcription factor that regulates developmental complexity in *Aspergillus nidulans*. *EMBO Journal*, **16:** 5710–5721.

DZIKOWSKA, A., KACPRZAK, M., TOMECKI, R., KOPER, M., SCAZZOCCHIO, C. & WEGLENSKI, P. (2003) Specific induction and carbon/nitrogen repression of arginine catabolism gene of *Aspergillus nidulans* – functional in vivo analysis of the *otaA* promoter. *Fungal Genetics & Biology*, **38:** 175–186.

EBEL, F., SCHWIENBACHER, M., BEYER, J., HEESEMANN, J., BRAKHAGE, A. A. & BROCK, M. (2006) Analysis of the regulation, expression, and localisation of the isocitrate lyase from *Aspergillus fumigatus*, a potential target for antifungal drug development. *Fungal Genetics & Biology*, **43:** 476–89.

EDWARDS, J., CHAMBERLAIN, D., BROSNAN, G., WEST, D., STANLEY, M. S., CLIPSON, N. J. W. & HOOLEY, P. (1998) A comparative physiological and morphological study of *Dendryphiella salina* and *D. arenaria* in relation to adaptation to life in the sea. *Mycological Research*, **102:** 1198–1202.

EHRLICH, K. C., MONTALBANO, B. G. & COTTY, P. J. (2003) Sequence comparison of *aflR* from different *Aspergillus* species provides evidence for variability in regulation of aflatoxin production. *Fungal Genetics & Biology*, **38:** 63–74.

EHRLICH, K. C., YU, J. & COTTY, P. J. (2005) Aflatoxin biosynthesis gene clusters and flanking regions. *Journal of Applied Microbiology*, **99:** 518–527.

EMPEL, J., SITKIEWICZ, I., ANDRUKIEWICZ A., LASOCKI, K., BORSUK, P. & WEGLENSKI, P. (2001) *arcA*, the regulatory gene for the arginine catabolic pathway in *Aspergillus nidulans*. *Molecular Genetics & Genomics*, **266:** 591–597.

ENJALBERT, B., NANTEL, A. & WHITEWAY, M. (2003) Stress-induced gene expression in *Candida albicans*: Absence of a general stress response. *Molecular Biology of the Cell*, **14:** 1460–1467.

ESPESO, E. A. & PEÑALVA, M. A. (1996) Three binding sites for the *Aspergillus nidulans* PacC zinc-finger transcription factor are necessary and sufficient for regulation by ambient pH of the isopenicillin N synthase gene promoter. *Journal of Bioloogical Chemistry*, **271:** 28825–28830.

ESPESO, E. A., TILBURN, J., SÁNCHEZ-PULIDO, L., BROWN, C.V., VALENCIA, A., ARST, H. N. & PENALVA, M. A. (1997) Specific DNA recognition by the *Aspergillus nidulans* three zinc finger transcription factor PacC *Journal of Molecular Biology*, **274:** 466–480.

ESPINEL-INGROFF, A. (2009) Novel antifungal agents, targets or therapeutic strategies for the treatment of invasive fungal diseases: a review of the literature (2005–2009). *Revista Iberoamericana de Micología*, **26:** 15–22.

ETXEBESTE, O., UGALDE, U. & ESPESO, E. A. (2010) Adaptative and developmental responses to stress in *Aspergillus nidulans*. *Current Protein & Peptide Science*, **11:** 704–118.

FAGUNDES, M. R. Z. K, FERNANDES, L., HARRIS, S. D., GOLDMAN, M. H. S. & GOLDMAN, G. H. (2003) Identification of a topoisomerase I mutant, *scsA1*, as an extragenic suppressor of a mutation in *scaA*^{NBS1}, the apparent homolog of human nibrin in *Aspergillus nidulans*. *Genetics*, **164:** 935–945

FAIRHEAD, C., LLORENTE, B., DENIS, F., SOLER, M. & DUJON, B. (1996) New vectors for combinatorial deletions in yeast chromosomes and for gap-repair cloning using 'split-marker' recombination. *Yeast*, **12:** 1439–1457.

FEKETE, E., KARAFFA, L., SÁNDOR, E., BÁNYAI, I., SEIBOTH, B., GYÉMÁNT, G., SEPSI, A., SZENTIRMAI, A. & KUBICEK, C. P. (2004) The alternative D-galactose degrading pathway of *Aspergillus nidulans* proceeds *via* L-sorbose. *Archives of Microbiology*, **181:** 35–44.

FELENBOK, B., FLIPPHI, M. & NIKOLAEV, I. (2001) Ethanol catabolism in *Aspergillus nidulans*: A model system for studying gene regulation. *Progress in Nucleic Acid Research*, **69:** 149–204.

FERNANDES, M., KELLER, N. P. & ADAMS, T. H. (1998) Sequence-specific binding by *Aspergillus nidulans* AflR, a C_6 zinc cluster protein regulating mycotoxin biosynthesis. *Molecular Microbiology*, **28:** 1355–1365.

FILLINGER, S., CHAVEROCHE, M. K., VAN DIJCK, P., DE VRIES, R., RUIJTER, G., THEVELEIN, J. & D'ENFERT, C. (2001) Trehalose is required for the acquisition of tolerance to a variety of stresses in the filamentous fungus *Aspergillus nidulans*. *Microbiology*, **147:** 1851–1862.

FINDON, H., CALCAGNO-PIZARELLI, A. M., MARTÍNEZ, J. L., SPIELVOGEL, A., MARKINA-IÑARRAIRAEGUI, A., INDRAKUMAR, T., RAMOS, J., PEÑALVA, M. A., ESPESO, E. A. & ARST, H. N. (2010) Analysis of a novel calcium autotrophy in *Aspergillus nidulans*. *Fungal Genetics & Biology*, **47:** 647–655.

FITZPATRICK, D. A., LOGUE, M. E., STAJICH, J. E. & BUTLER, G. (2006) A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evolutionary Biology*, **6:** 99.

FLECK, C. B., SCHÖBEL, F. & BROCK, M. (2011) Nutrient acquisition by pathogenic fungi: Nutrient availability, pathway regulation, and differences in substrate utilization. *International Journal of Medical Microbiology*, **301:** 400–407.

FLIPPHI, M., KOCIALKOWSKA, J. & FELENBOK, B. (2002) Characteristics of physiological inducers of the ethanol utilization (alc) pathway in *Aspergillus nidulans*. *Biochemical Journal*, **364:** 25–31.

FLIPPHI, M., SUN, J., ROBELLET, X., KARAFFA, L., FEKETE, E., ZENG, A. KUBICEK, C. P. (2009) Biodiversity and evolution of primary carbon metabolism in *Aspergillus nidulans* and other *Aspergillus* spp. *Fungal Genetics & Biology*, **46:** S19–S44.

FRITH, M. C., HANSEN, U., SPOUGE, J. L. & WENG, Z. (2004) Finding functional sequence elements by multiple local alignment. *Nucleic Acids Research*, **32:** 189–200.

FUJIOKA, T., MIZUTANI, O., FURUKAWA, K., SATO, N., YOSHIMI, A., YAMAGATA, Y., NAKAJIMA, T. & ABE, K. (2007) MpkA-dependent and -independent cell wall integrity signaling in *Aspergillus nidulans*. *Eukaryotic Cell*, **6:** 1497–1510.

FURUKAWA, K., HOSHI, Y., MAEDA ,T., NAKAJIMA, T. & ABE, K. (2005) *Aspergillus nidulans* HOG pathway is activated only by two-component signalling pathway in response to osmotic stress. *Molecular Microbiology*, **56:** 1246−1261.

FUTAGAMI, T., NAKAO, S., KIDO, Y., OKA, T., KAJIWARA, Y., TAKASHITA, H., OMORI, T., FURUKAWA, K. & GOTO, M. (2011) Putative stress sensors WscA and WscB are involved in hypo-osmotic and acidic pH stress tolerance in *Aspergillus nidulans*. *Eukaryotic Cell*, **10:** 1504–1515.

GALAGAN, J. E., CALVO, S. E., CUOMO, C., MA, L., WORTMAN, J., BATZOGLOU, S., LEE, S., BRUDNO, M., BASTURKMEN, M., SPEVAK, C. C., CLUTTERBUCK, J., KAPITONOV, V., JURKA, J., SCAZZOCCHIO, C., FARMAN, M., BUTLER, J., PURCELL, S., HARRIS, S., BRAUS, G. H., DRAHT, O., BUSCH, S., D'ENFERT, C., BOUCHIER, C., GOLDMAN, G. H., BELL-PEDERSEN, D., GRIFFITHS-JONES, S., DOONAN, J. H., YU, J., VIENKEN, K., PAIN, A., FREITAG, M., SELKER, E. U., ARCHER, D. B., PEÑALVA, M. A., OAKLEY, B. R., MOMANY, M., TANAKA, T., KUMAGAI, T., ASAI, K., MACHIDA, M., NIERMAN, W. C., DENNING, D. W., CADDICK, M., HYNES, M., PAOLETTI, M., FISCHER, R., MILLER, B., DYER, P., SACHS, M. S., OSMANI, S. A. & BIRREN, B. W. (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae. Nature*, **438:** 1105–1115.

GAMSJAEGER, R., LIEW, C. K., LOUGHLIN, F. E., CROSSLEY, M. & MACKAY, J. P. (2007) Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends in Biochemical Sciences*, **32:** 63–70.

GANCEDO, J. M. (1998) Yeast carbon catabolite repression. *Microbiology & Molecular Biology Reviews*, **62:** 334–361.

GARCÍA, I., GONZALEZ, R., GÓMEZ, D. & SCAZZOCCHIO, C. (2004) Chromatin rearrangements in the *prnD-prnB* bi-directional promoter: dependence on transcription factors. *Eukaryotic Cell*, **3:** 144–156.

GASCH, A. P. (2007) Comparative genomics of the environmental stress response in ascomycete fungi. *Yeast*, **24:** 961–976.

GASCH, A. P., SPELLMAN, P. T., KAO, C. M., CARMEL-HAREL, O., EISEN, M. B., STORZ, G., BOTSTEIN, D. & BROWN, P. O. (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Molecular Biology of the Cell*, **11:** 4241–4257.

GAUWERKY, K. BORELLI, C. & KORTING, H. C. (2009) Targeting virulence: A new paradigm for antifungals. *Drug Discovery Today*, **14:** 214–222.

GOINS, C. L., GERIK, K. J. & LODGE, J. K. (2006) Improvements to gene deletion in the fungal pathogen *Cryptococcus neoformans*: absence of Ku proteins increases homologous recombination, and co-transformation of independent DNA molecules allows rapid complementation of deletion phenotypes. *Fungal Genetics & Biology*, **43:** 531–544.

GOLDMAN, G. H. & KAFER, E. (2004) *Aspergillus nidulans* as a model system to characterize the DNA damage response in eukaryotes. *Fungal Genetics & Biology*, **41:** 428–442.

GOLDMAN, G. H., McGUIRE, S. L. & HARRIS, S. D. (2002) The DNA damage response in filamentous fungi. *Fungal Genetics & Biology*, **35:** 183–195.

GÓMEZ, D., GARCÍA, I., SCAZZOCCHIO, C. & CUBERO, B. (2003) Multiple GATA sites: protein binding and physiological relevance for the regulation of the proline transporter gene of *Aspergillus nidulans. Molecular Microbiology*, **50:** 277–289.

GONZALEZ, R., GAVRIAS, V., GOMEZ, D., SCAZZOCCHIO, C. & CUBERO, B. (1997) The integration of nitrogen and carbon catabolite repression in *Aspergillus nidulans* requires the GATA factor AreA and an additional positive-acting element, ADA. *The EMBO Journal*, **16:** 2937–2944.

GREMEL, G., DORRER, M. & SCHMOLL, M. (2008) Sulphur metabolism and cellulase gene expression are connected processes in the filamentous fungus *Hypocrea jecorina* (anamorph *Trichoderma reesei*). *BMC Microbiology*, **8:** 14.

GUEST, G. M., LIN, X. & MOMANY, M. (2004) *Aspergillus nidulans* RhoA is involved in polar growth, branching, and cell wall synthesis. *Fungal Genetics & Biology*, **41:** 13–22.

GURURANI, M. A., VENKATESH, J., UPADHYAYA, C. P., NOOKARAJU, A., PANDEY, S. K. & PARK, S. W. (2012) Plant disease resistance genes: Current status and future directions. *Physiological and Molecular Plant Pathology*, **78:** 51–65.

GUO, Y. & LEVIN, H. L. (2010) High-throughput sequencing of retrotransposon integration provides a saturated profile of target activity in *Schizosaccharomyces pombe*. *Genome Research*, **20:** 239–248.

GUSTIN, M. C., ALBERTYN, J., ALEXANDER, M. & DAVENPORT, K. (1998) MAP kinase pathways in the yeast *Saccharomyces cerevisiae*. *Microbiology & Molecular Biology Reviews*, **62:** 1264–1300.

HAAS, B. J., ZENG, Q., PEARSON, M. D., CUOMO, C. A. & WORTMAN, J. R. (2011) Approaches to fungal genome annotation. *Mycology*, **2:** 118−141.

HAGIWARA, D., ASANO, Y., MARUI, J., YOSHIMI, A., MIZUNO, T. & ABE, K. (2009) Transcriptional profiling for *Aspergillus nidulans* HogA MAPK signaling pathway in response to fludioxonil and osmotic stress. *Fungal Genetics & Biology*, **46:** 868–878.

HALLSWORTH, J. E., PRIOR, B. A., NOMURA, Y., IWAHARA, M. & TIMMIS, K, N. (2003) Compatible solutes protect against chaotrope (ethanol)-induced, nonosmotic water stress. *Applied and Environmental Microbiology*, **69:** 7032−7034.

HAN, K. H. & PRADE, R. A. (2002) Osmotic stress-coupled maintenance of polar growth in *Aspergillus* nidulans. *Molecular Microbiology*, **43:** 1065–1078.

HAN, K. H., SEO, J. A. & YU, J. H. (2004) Regulators of G-protein signalling in *Aspergillus nidulans*: RgsA downregulates stress response and stimulates asexual sporulation through attenuation of GanB (Gα) signalling. *Molecular Microbiology*, **53:** 529–540.

HANAHAN, D. (1985) In *DNA Cloning: A Practical Approach* (Ed, Glover, D.). Vol. 1, 109–135. IRL Press Ltd, London.

HANSBERG, W., SALAS-LIZANA, R. & DOMÍNGUEZ, L. (2012) Fungal catalases: function, phylogenetic origin and structure. *Archives of Biochemistry and Biophysics*, **525:** 170–180.

HAO, Z., FURUNOBU, A., NAGATA, A. & OKAYAMA, H. (1997) A zinc finger protein required for stationary phase viability in fission yeast. *Journal of Cell Science*, **110:** 2557–2566.

HARRIS, S. D., TURNER, G., MEYER, V., ESPESO, E. A., SPECHT, T., TAKESHITA, N. & HELMSTEDT, K. (2009) Morphology and development in *Aspergillus nidulans*: A complex puzzle. *Fungal Genetics & Biology*, **46:** S82–S92.

HARTMANN, T., SASSE, C., SCHEDLER, A., HASENBERG, M., GUNZER, M. & KRAPPMANN, S. (2011) Shaping the fungal adaptome – Stress responses of *Aspergillus fumigatus*. *International Journal of Medical Microbiology*, **301:** 408–416.

HATAKEYAMA, S., ITO, Y., SHIMANE, A., ISHII, C. & INOUE, H. (1998) Cloning and characterization of the yeast *RAD1* homolog gene (*mus-38*) from *Neurospora crassa*: evidence for involvement in nucleotide excision repair. *Current Genetics*, **33:** 276–283.

HIBBETT, D. S., BINDER, M., BISCHOFF, J. F., BLACKWELL, M., CANNON, P. F., ERIKSSON, O. E., HUHNDORF, S., JAMES, T., KIRK, P. M., LÜCKING, R., THORSTEN-LUMBSCH, H, LUTZONI, F, MATHENY, P. B, McLAUGHLIN, D. J., POWELL, M. J., REDHEAD, S., SCHOCH, C. L., SPATAFORA, J. W., STALPERS, J. A., VILGALYS, R., AIME, M. C., APTROOT, A., BAUER, R., BEGEROW, D., BENNY, G. L., CASTLEBURY, L. A., CROUS, P. W., DAI, Y. C., GAMS, W., GEISER, D. M., GRIFFITH, G. W., GUEIDAN, C., HAWKSWORTH, D. L., HESTMARK, G., HOSAKA, K., HUMBER, R. A., HYDE, K. D., IRONSIDE, J. E., KÕLJALG, U., KURTZMAN, C. P., LARSSON, K. H., LICHTWARDT, R., LONGCORE, J., MIADLIKOWSKA, J., MILLER, A., MONCALVO, J. M., MOZLEY-STANDRIDGE, S., OBERWINKLER, F., PARMASTO, E., REEB, V., ROGERS, J. D., ROUX, C., RYVARDEN, L., SAMPAIO, J. P., SCHÜSSLER, A., SUGIYAMA, J., THORN, R. G., TIBELL, L., UNTEREINER, W. A., WALKER, C., WANG, Z., WEIR, A., WEISS, M., WHITE, M. M., WINKA, K., YAO, Y. J. & ZHANG, N. (2007) A higher-level phylogenetic classification of the Fungi. *Mycological Research*, **111:** 509–547.

HINNEBUSCH, A. G. & NATARAJAN, K. (2002) Gcn4p, a master regulator of gene expression, is controlled at multiple levels by diverse signals of starvation and stress. *Eukaryotic Cell*, **1:** 22–32.

HIRASAWA, T. & YOKOIGAWA, K. (2001) Leavening ability of baker's yeast exposed to hyperosmotic media. *FEMS Microbiology Letters*, **194:** 159–162.

HIRASAWA, T., ASHITANI, K., YOSHIKAWA, K., NAGAHISA, K., FURUSAWA, C., KATAKURA, Y., SHIMIZU, H. & SHIOYA, S. (2006) Comparison of transcriptional responses to osmotic stresses induced by NaCl and sorbitol additions in *Saccharomyces cerevisiae* using DNA microarray. *Journal of Bioscience and Bioengineering*, **102:** 568–571.

HOFFMANN, A. F. & HARRIS, S. D. (2001) The *Aspergillus nidulans musN* gene encodes a RecQ helicase that interacts with the PI-3K-related kinase *uvsB*. *Genetics*, **159:** 1595–1604.

HOFFMANN, B., WANKE, C., LAPAGLIA, S. K. & BRAUS, G. H. (2000) c-Jun and RACK1 homologues regulate a control point for sexual development in *Aspergillus nidulans*. *Molecular Microbiology*, **37:** 28–41.

HOFFMANN, B., VALERIUS, O., ANDERMANN, M. & BRAUS, G. H. (2001) Transcriptional autoregulation and inhibition of mRNA translation of amino acid regulator gene *cpcA* of filamentous fungus *Aspergillus nidulans*. *Molecular Biology of the Cell*, **12:** 2846–2857.

HOHMANN, S. (2009) Control of high osmolarity signalling in the yeast *Saccharomyces cerevisiae*. *FEBS Letters*, **583:** 4025–4029.

HOOLEY, P., SHAWCROSS, S. G. & STRIKE, P. (1988) Mutants of *Aspergillus nidulans* with increased resistance to the alkylating agent *N*-Methyl-*N*'-Nitro-*N*-Nitrosoguanidine. *Mutation Research*, **193:** 247–254.

HYNES, M. J., DRAHT, O. W. & DAVIS, M. A. (2002) Regulation of the *acuF* gene, encoding phosphoenolpyruvate carboxykinase in the filamentous fungus *Aspergillus nidulans*. *Journal of Bacteriology*, **184:** 183–190.

HYNES, M. J. MURRAY, S. L. DUNCAN, A. KHEW, G. S. & DAVIS, M. A. (2006) Regulatory genes controlling fatty acid catabolism and peroxisomal functions in the filamentous fungus *Aspergillus nidulans*. *Eukaryotic Cell*, **5:** 794–805.

HYNES, M. J., SZEWCZYK, E., MURRAY, S. L., SUZUKI, Y., DAVIS, M. A. & SEALY-LEWIS, H. M. (2007) Transcriptional control of gluconeogenesis in *Aspergillus nidulans*. *Genetics*, **176:** 139–150.

IDNURM, A. & HOWLETT, B. J. (2002) Isocitrate lyase is essential for pathogenicity of the fungus *Leptosphaeria maculans* to canola (*Brassica napus*). *Eukaryotic Cell*, **1:** 719– 724.

INOUE, H., NOJIMA, H. & OKAYAMA, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, **96:** 23–28.

IUCHI, S. (2001) Three classes of C_2H_2 zinc finger proteins. *Cellular and Molecular Life Sciences*, **58:** 625–635.

JACOBS, G. H. (1992) Determination of the base recognition positions of zinc fingers from sequence analysis. *The EMBO Journal*, **11:** 4507–4517.

JEFFARES, D. C., MOURIER, T. & PENNY, D. (2006) The biology of intron gain and loss. *Trends in Genetics*, **22:** 16–22.

JIN, Y., BOK, J. W., GUZMAN-DE-PEÑA, D. & KELLER, N. P. (2002) Requirement of spermidine for developmental transitions in *Aspergillus nidulans*. *Molecular Microbiology*, **46:** 801–812.

JONES, D. T. (1999) GenTHREADER: An efficient and reliable protein fold recognition method for genomic sequences. *Journal of Molecular Biology*, **287:** 797–815.

JONES, G. W., HOOLEY, P., FARRINGTON, S. M., SHAWCROSS, S. G., IWANEJKO, L. A. & STRIKE, P. (1999) Cloning and characterisation of the *sagA* gene of *Aspergillus nidulans*: a gene which affects sensitivity to DNA-damaging agents. *Molecular & General Genetics*, **261:** 251–258.

KANEKO, I., TANAKA, A. & TSUGE, T. (2000) *REAL*, an LTR retrotransposon from the plant pathogenic fungus *Alternaria alternata*. *Molecular & General Genetics*, **263:** 625–634.

KAWASAKI, L. & AGUIRRE, J. (2001) Multiple catalase genes are differentially regulated in *Aspergillus nidulans*. *Journal of Bacteriology*, **183:** 1434−1440.

KELLY, J. M. & BOASE, N. A. (2004) A role for *creD*, a carbon catabolite repression gene from *Aspergillus nidulans*, in ubiquitination. *Molecular Microbiology*, **53:** 929−940.

KNIGHT, R. D. & SHIMELD, S. M. (2001) Identification of conserved C2H2 zinc-finger gene families in the Bilateria. *Genome Biology,* **2 (5):** 0016.1–0016.8.

KOPECKÁ, M. & GABRIEL, M. (1992) The influence of congo red on the cell wall and (1→3)-beta-D-glucan microfibril biogenesis in *Saccharomyces cerevisiae*. *Archives of Microbiology*, **158:** 115–126.

KRAPPMANN, S. (2007) Gene targeting in filamentous fungi: the benefits of impaired repair. *Fungal Biology Reviews*, **21:** 25–29.

KRAPPMANN, S. & BRAUS, G. H. (2005) Nitrogen metabolism of *Aspergillus* and its role in pathogenicity. *Medical Mycology*, **43:** S31–S40.

KRAPPMANN, S., BIGNALL, E. M., UTZ, R., ROGERS, T., HAYNES, K. & BRAUS, G. H. (2004) The *Aspergillus fumigatus* transcriptional activator CpcA contributes significantly to the virulence of this fungal pathogen. *Molecular Microbiology*, **52:** 785–799.

KRAPPMANN, S., SASSE, C. & BRAUS, G. H. (2006) Gene targeting in *Aspergillus fumigatus* by homologous recombination is facilitated in a non-homologous end-joining-deficient genetic background. *Eukaryotic Cell*, **5:** 212–215.

KRAUS, P. R. & HARRIS, S. D. (2001) The *Aspergillus nidulans snt* genes are required for the regulation of septum formation and cell cycle checkpoints. *Genetics*, **159:** 557–569.

KRIZKOVA, B. (2010) *Characterisation of the sltA1 mutation in Aspergillus nidulans*. B.Sc. Honours Project, University of Wolverhampton.

KUBICEK, C. P., MIKUS, M., SCHUSTER, A., SCHMOLL, M. & SEIBOTH, B. (2009) Metabolic engineering strategies for the improvement of cellulase production by *Hypocrea jecorina*. *Biotechnology for Biofuels*, **2:** 19.

KÜCK, U. & HOFF, B. (2010) New tools for the genetic manipulation of filamentous fungi. *Applied Microbiology & Biotechnology*, **86:** 51–62.

KUDLA, B., CADDICK, M. X., LANGDON, T., MARTINEZ-ROSSI, N. M., BENNETT, C. F., SIBLEY, S., DAVIES, R. W. & ARST, H. N. (1990) The regulatory gene *areA* mediating nitrogen metabolite repression in *Aspergillus nidulans*. Mutations affecting specificity of gene activation alter a loop residue of a putative zinc finger. *EMBO Journal*, **9:** 1353–1364.

KULMBURG, P., MATHIEU, M., DOWZER, C., KELLY, J. & FELENBOK, B. (1993) Specific binding sites in the *alcR* and *alcA* promoters of the ethanol regulon for the CREA repressor mediating carbon catabolite repression in *Aspergillus nidulans*. *Molecular Microbiology*, **7:** 847–857.

KUMAR, L., BREAKSPEAR, A., KISTLER, C., MA, L. J. & XIE, X. (2010) Systematic discovery of regulatory motifs in *Fusarium graminearum* by comparing four *Fusarium* genomes. *BMC Genomics*, **11:** 208.

KUPFER, D. M., DRABENSTOT, S. D., BUCHANAN, K. L., LAI, H., ZHU, H., DYER, D. W., ROE, B. A. & MURPHY, J. W. (2004) Introns and splicing elements of five diverse fungi. *Eukaryotic Cell*, **3:** 1088–1100.

LAITY, J. H., LEE, B. M. & WRIGHT, P. E. (2001) Zinc finger proteins: new insights into structural and functional diversity. *Current Opinion in Structural Biology*, **11:** 39–46.

LARABEE, J. L., HOCKER, J. R. & HANAS, J. S. (2005) Cys redox reactions and metal binding of a Cys_2 His_2 zinc finger. *Archives of Biochemistry and Biophysics*, **434:** 139–149.

LARA-ROJAS, F., SÁNCHEZ, O., KAWASAKI, L. & AGUIRRE, J. (2011) *Aspergillus nidulans* transcription factor AtfA interacts with the MAPK SakA to regulate general stress responses, development and spore functions. *Molecular Microbiology*, **80:** 436–454.

LARRONDO, L. F., COLOT, H. V., BAKER, C. L., LOROS, J. J. & DUNLAP, J. C. (2009) Fungal functional genomics: tunable knockout strategies – knock-in expression and tagging strategies. *Eukaryotic Cell*, **8:** 800–804.

LATCHMAN, D. S. (1998) *Eukaryotic Transcription Factors*. 3rd edition. Academic Press, London.

LEVIN, D. E. (2011) Regulation of cell wall biogenesis in *Saccharomyces cerevisiae*: the cell wall integrity signaling pathway. *Genetics*, **189:** 1145–1175.

LEWIS, J. G., LEARMONTH, R. P. & WATSON. K. (1995) Induction of heat, freezing and salt tolerance by heat and salt shock in *Saccharomyces cerevisiae*. *Microbiology*, **141:** 687–694.

LI, Q., HARVEY, L. M. & McNEIL, B. (2009) Oxidative stress in industrial fungi. *Critical Reviews in Biotechnology*, **29:** 199–213.

LIU, Y. J. & HALL, B. D. (2004) Body plan evolution of ascomycetes, as inferred from an RNA polymerase II phylogeny. *Proceedings of the National Academy of Sciences*, **101:** 4507–4512.

LOCKINGTON, R. & KELLY, J. (2002) The WD40-repeat protein CreC interacts with and stabilizes the deubiquitinating enzyme CreB *in vivo* in *Aspergillus nidulans*. *Molecular Microbiology*, **43:** 1173–1182.

LOCKINGTON, R. A., RODBOURN, L., BARNETT, S., CARTER, C. J. & KELLY, J. M. (2002) Regulation by carbon and nitrogen sources of a family of cellulases in *Aspergillus nidulans*. *Fungal Genetics & Biology*, **37:** 190–196.

LORANG, J. M., TUORI, R. P., MARTÍNEZ, J. P., SAWYER, T. L., REDMAN, R. S., ROLLINS, J. A., WOLPERT, T. J., JOHNSON, K. B., RODRIGUEZ, R. J., DICKMAN, M. B. & CIUFFETTI, L. M. (2001) Green fluorescent protein is lighting up fungal biology. *Applied & Environmental Microbiology*, **67:** 1987–1994.

LORENZ, M. C. & FINK, G. R. (2001) The glyoxylate cycle is required for fungal virulence. *Nature*, **412:** 83–86.

LUTZONI, F., KAUFF, F., COX, C. J., McLAUGHLIN, D., CELIO, G., DENTINGER, B., PADAMSEE, M., HIBBETT, D., JAMES, T. Y., BALOCH, E., GRUBE, M., REEB, V., HOFSTETTER, V., SCHOCH, C., ARNOLD, A. E., MIADLIKOWSKA, J., SPATAFORA, J., JOHNSON, D., HAMBLETON, S., CROCKETT, M., SHOEMAKER, R., SUNG, G. H., LÜCKING, R., LUMBSCH, T., O'DONNELL, K., BINDER, M., DIEDERICH, P., ERTZ, D., GUEIDAN, C., HANSEN, K., HARRIS, R. C, HOSAKA, K., LIM, Y. W., MATHENY, B.,

NISHIDA, H., PFISTER, D., ROGERS, J., ROSSMAN, A., SCHMITT, I., SIPMAN, H., STONE, J., SUGIYAMA, J., YAHR, R. & VILGALYSM, R. (2004) Assembling the fungal tree of life: Progress, classification, and evolution of subcellular traits. *American Journal of Botany*, **91:** 1446–1480.

MABEY, J. E., ANDERSON, M. J., GILES, P. F., MILLER, C. J., ATTWOOD, T. K., PATON, N. W., BORNBERG-BAUER, E., ROBSON, G. D., OLIVER, S. G. & DENNING, D. W. (2004) C*A*DRE: the Central *Aspergillus* Data REpository. *Nucleic Acids Research*, **32:** D401–D405.

MACHIDA, M. & GOMI, K. *Aspergillus: Molecular Biology and Genomics*. (2010) Caister Academic Press, Norfolk, UK.

MACIOS, M., CADDICK, M. X., WEGLENSKI, P., SCAZZOCCHIO, C. & DZIKOWSKA, A. (2012) The GATA factors AREA and AREB together with the co-repressor NMRA, negatively regulate arginine catabolism in *Aspergillus nidulans* in response to nitrogen and carbon source. *Fungal Genetics & Biology*, **49:** 189–198.

MAGER, W. H. & DE KRUIJFF, A. J. (1995) Stress-induced transcriptional activation. *Microbiology Reviews*, **59:** 506–531.

MALAVAZI, I., SAVOLDI, M., MAURO, S. M. Z., MENCK, C. F. M., HARRIS, S. D., GOLDMAN, M. H. & GOLDMAN, G. H. (2006) Transcriptome analysis of *Aspergillus nidulans* exposed to captothecin-induced DNA damage. *Eukaryotic* Cell, **5:** 1688–1704.

MANNUSS, A., TRAPP, O. & PUCHTA, H. (2012) Gene regulation in response to DNA damage. *Biochimica et Biophysica Acta*, **1819:** 154–165.

MAQUAT, L. E. (2005) Nonsense-mediated mRNA decay in mammals. *Journal of Cell Science*, **118:** 1773–1776.

MARSDEN, R. L. McGUFFIN, L. J. & JONES, D. T. (2002) Rapid protein domain assignment from amino acid sequence using predicted secondary structure. *Protein Science*, **11:** 2814–2824.

MARTÍNEZ-PASTOR, M. T., MARCHLER, G., SCHÜLLER, C., MARCHLER-BAUER, A., RUIS, H. & ESTRUCH, F. (1996) The *Saccharomyces cerevisiae* zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). *EMBO Journal*, **15:** 2227–2235.

MARZLUF, G. A. (1997) Genetic regulation of nitrogen metabolism in the fungi. *Microbiology and Molecular Biology Reviews*, **61:** 17–32.

McCAHILL, A., WARWICKER, J., BOLGER, G. B., HOUSLAY, M. D. & YARWOOD, S. J. (2002) The RACK1 scaffold protein: a dynamic cog in cell response systems. *Molecular Pharmacology*, **62:** 1261–1273.

McGUFFIN, L. J., BRYSON, K. & JONES, D. T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics*, **16:** 404–405.

McGUIRE, A. M., PEARSON, M. D., NEAFSEY, D. E. & GALAGAN, J. E. (2008) Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biology*, **9:** R50.

MILLER, G. L. (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Analytical Chemistry*, **31:** 426–428.

MISKEI, M., KARÁNYI, Z. & PÓCSI, I. (2009) Annotation of stress-response proteins in the *Aspergilli*. *Fungal Genetics & Biology*, **46:** (1) S105–120.

MOROZOV, I. Y., NEGRETTE-URTASUN, S., TILBURN, J., JANSEN, C. A., CADDICK, M. X., ARST, H. N. (2006) Nonsense-mediated mRNA decay mutation in *Aspergillus nidulans*. *Eukaryotic Cell*, **5:** 1838–1846.

MORSCHHÄUSER, J. (2010) Regulation of multidrug resistance in pathogenic fungi. *Fungal Genetics & Biology*, **47:** 94–106.

MULDER, H. J., NIKOLAEV, I. & MADRID, S. M. (2006) HacA, the transcriptional activator of the unfolded protein response (UPR) in *Aspergillus niger*, binds partly to palindromic UPR elements of the consensus sequence 5'-CAN(G/A)NTGT/GCCT-3'. *Fungal Genetics & Biology*, **43:** 560–572.

MULLIS, K., FALOONA, F., SCHARF, S., SAIKI, R., HORN, G. & ERLICH, H. (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology*, **51:** 263–273.

MURPHY, R. L., ANDRIANOPOULOS, A., DAVIS, M. A. & HYNES, M. J. (1997) Identification of *amdX*, a new Cys-2-His-2 (C2H2) zinc-finger gene involved in the regulation of the *amdS* gene of *Aspergillus nidulans*. *Molecular Microbiology*, **23:** 591–602.

NASS, R. & RAO, R. (1999) The yeast endosomal Na^+/H^+ exchanger, Nhx1, confers osmotolerance following acute hypertonic shock. *Microbiology*, **145:** 3221–3228.

NATARAJAN, K., MEYER, M. R., JACKSON, B. M., SLADE, D., ROBERTS, C., HINNEBUSCH, A. G. & MARTON, M. J. (2001) Transcriptional profiling shows Gcn4p is a master regulator of gene expression during amino acid starvation in yeast. *Molecular & Cellular Biology*, **21:** 4347–4368.

NAYAK, T., SZEWCZYK, E., OAKLEY, C. E., OSMANI, A., UKIL, L., MURRAY, S. L., HYNES, M. J., OSMANI, S. A. & OAKLEY, B. R. (2006) A versatile and efficient gene-targeting system for *Aspergillus nidulans*. *Genetics*, **172:** 1557–1566.

NEGRETE-URTASUN, S., DENISON, S. H. & ARST, H. N. (1997) Characterization of the pH signal transduction pathway gene *palA* of *Aspergillus nidulans* and identification of possible homologs. *Journal of Bacteriology*, **179:** 1832–1835.

NEWBURY, J. & PEBERDY, J. F. (1996) Characterization of the heat shock response in protoplasts of *Aspergillus nidulans*. *Mycological Research*, **100:** 1325–1332.

NIELSEN, C. B., FRIEDMAN, B., BIRREN, B., BURGE, C. B. & GALAGAN, J. E. (2004) Patterns of intron loss and gain in fungi. *PLoS Biology*, **2:** e422.

NIELSEN, M. L., ALBERTSON, L., LETTIER, G., NIELSON, J. B. & MORTENSEN, U. H. (2006) Efficient PCR-based gene targeting with a recyclable marker in *Aspergillus nidulans*. *Fungal Genetics & Biology*, **43:** 54–64.

NIELSON, J. B., NIELSEN, M. L. & MORTENSEN, U. H. (2008) Transient disruption of non-homologous end-joining facilitates targeted genome manipulations in the filamentous fungus *Aspergillus nidulans*. *Fungal Genetics & Biology*, **45:** 165–170.

NIGG, E. A. (1997) Nucleocytoplasmic transport: signals, mechanisms and regulation. *Nature*, **386:** 779–787.

NIKOLAOU, E., AGRAFIOTI, I., STUMPF, M., QUINN, J., STANSFIELD, I. & BROWN, A. J. P. (2009) Phylogenetic diversity of stress signalling pathways in fungi. *BMC Evolutionary Biology*, **9:** 44.

NINOMIYA, Y., SUZUKI, K., ISHII, C. & INOUE, H. (2004) Highly efficient gene replacements in *Neurospora* strains deficient for non-homologous end-joining. *Proceedings of the National Academy of Sciences*, **101:** 12248–12253.

NOTREDAME, C., HIGGINS, D. G. & HERINGA, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, **302:** 205–217.

NOVIKOVA, O. S., FET, V. & BLINOV, A. G. (2007) LTR retrotransposons in the *Aspergillus fumigatus* and *A. nidulans* genomes. *Molecular Biology*, **41:** 756–763.

OAKLEY, B. R., RINEHART, J. E., MITCHELL, B. L., OAKLEY, C. E., CARMONA, C., GRAY, G. L. & MAY, G. S. (1987) Cloning, mapping and molecular analysis of the *pyrG* (orotidine-5'-phosphate decarboxylase) gene of *Aspergillus nidulans*. *Gene*, **61:** 385–399.

O'HANLON, K. A., MARGISON, G. P., HATCH, A., FITZPATRICK, D. A., OWENS, R. A., DOYLE, S. & JONES, G. W. (2012) Molecular characterization of an adaptive response to alkylating agents in the opportunistic pathogen *Aspergillus fumigatus*. *Nucleic Acids Research*, **40:** 7806–7820.

OLIVAS, I., ROYUELA, M., ROMERO, B., MONTEIRO, M. C., MINGUEZ, J. M., LABORDA, F. & DE LUCAS, J. R. (2008) Ability to grow on lipids accounts for the fully virulent phenotype in neutropenic mice of *Aspergillus fumigatus* null mutants in the key glyoxylate cycle enzymes. *Fungal Genetics & Biology*, **45:** 45–60.

OLSZEWSKA, A., KRÓL, K., WEGLENSKI, P. & DZIKOWSKA, A. (2007) Arginine catabolism in *Aspergillus nidulans* is regulated by the *rrmA* gene coding for the RNA-binding protein. *Fungal Genetics & Biology*, **44:** 1285–1297.

O'MAHONY, R. J., BURNS, A. T. H., MILLAM, S., HOOLEY, P. & FINCHAM, D. A. (2002) Isotopic growth of spores and salt tolerance in *Aspergillus nidulans*. *Mycological Research*, **106:** 1480−1486.

O'NEIL, J. D., BUGNO, M., STANLEY, M. S., BARHAM-MORRIS, J. B., WOODCOCK, N. A., CLEMENT, D. J., CLIPSON, N. J. W., WHITEHEAD, M. P., FINCHAM, D. A. & HOOLEY, P. H. (2002) Cloning of a novel gene encoding a C_2H_2 zinc finger protein that alleviates sensitivity to abiotic stresses in *Aspergillus nidulans*. *Mycological Research*, **106:** 491–498.

O'ROURKE, S. M. & HERSKOWITZ, I. (2004) Unique and redundant roles for HOG MAPK pathway components as revealed by whole genome expression analysis. *Molecular Biology of the Cell*, **15:** 532–542.

PANOZZO, C., CAPUANO, V., FILLINGER, S. & FELENBOK, B. (1997) The zinc binuclear cluster activator AlcR is able to bind to single sites but requires multiple repeated sites for synergistic activation of the *alcA* gene in *Aspergillus nidulans*. *Journal of Biological Chemistry*, **272:** 22859–22865.

PARMAR, J. H., BHARTIYA, S. & VENKATESH, K. V. (2009) A model-based study delineating the roles of the two signaling branches of *Saccharomyces cerevisiae*, Sho1 and Sln1, during adaptation to osmotic stress. *Physical Biology*, **6:** 036019.

PATERSON, R. R. M. & LIMA, N. (2008) Mutagens manufactured in fungal culture may affect DNA/RNA of producing fungi. *Journal of Applied Microbiology*, **106:** 1070–1080.

PEÑALVA, M. A., TILBURN, J., BIGNELL, E. & ARST, H. N. (2008) Ambient pH gene regulation in fungi: making connections. *Trends in Microbiology*, **16:** 291–300.

PERI, S., IBARROLA, N., BLAGOEV, B., MANN, M. & PANDEY, A. (2001) Common pitfalls in bioinformatics based analyses: look before you leap. *Trends in Genetics*, **17:** 541–545.

PETERSEN, K. L., LEHMBECK, J. & CHRISTENSEN, T. (1999) A new transcriptional activator for amylase genes in *Aspergillus*. *Molecular & General Genetics*, **262:** 668–676.

PINCHAI, N., JUVVADI, P. R., FORTWENDEL, J. R., PERFECT, B. Z., ROGG, L. E., ASFAW, Y. G. & STEINBACK, W. J. (2010) The *Aspergillus fumigatus* P-type Golgi apparatus Ca^{2+}/Mn^{2+} ATPase PmrA is involved in cation homeostasis and cell wall integrity but is not essential for pathogenesis. *Eukaryotic Cell*, **9:** 472–476.

PÖGGELER, S. & KÜCK, U. (2006) Highly efficient generation of signal transduction knockout mutants using a fungal strain deficient in the mammalian *ku70* ortholog. *Gene*, **378:** 1–10.

PONTECORVO, G., ROPER, J. A., HEMMONS, L. M., MACDONALD, K. D. & BUFTON, A. W. J. (1953) The genetics of *Aspergillus nidulans*. *Advanced Genetics*, **5:** 141–238.

QI, M. & ELION, E. A. (2005) MAP kinase pathways. *Journal of Cell Science*, **118:** 3569–3572.

RAEDER, U. & BRODA, P. (1985) Rapid preparation of DNA from filamentous fungi. *Letters in Applied Microbiology*, **1:** 17–20.

RAM, A. F., WOLTERS, A., TEN HOOPEN, R. & KLIS, F. M. (1994) A new approach for isolating cell wall mutants in *Saccharomyces cerevisiae* by screening for hypersensitivity to calcofluor white. *Yeast*, **10:** 1019–1030.

RATHS, S., ROHRER, J., CRAUSAZ, F. & RIEZMAN, H. (1993) *end3* and *end4*: two mutants defective in receptor-mediated and fluid-phase endocytosis in *Saccharomyces cerevisiae*. *The Journal of Cell Biology*, **120:** 55–65.

RAVAGNANI, A., GORFINKIEL, L., LANGDON, T., DIALLINAS, G., ADJADJ, E., DEMAIS, S., GORTON. D., ARST, H. N. & SCAZZOCCHIO, C. (1997). Subtle hydrophobic interactions between the seventh residue of the zinc finger loop and the first base of an HGATAR sequence determine promoter-specific recognition by the *Aspergillus nidulans* GATA factor AreA. *The EMBO Journal*, **16:** 1974–1986.

REDKAR, R. J., LOCY, R. D. & SINGH, N. K. (1995) Biosynthetic pathways of glycerol accumulation under salt stress in *Aspergillus nidulans*. *Experimental Mycology* **19:** 241–246.

REP, M., KRANTZ, M., THEVELEIN, J. M. & HOHMANN, S. (2000) The transcriptional response of *Saccharomyces cerevisiae* to osmotic shock. Hot1p and Msn2p/Msn4p are required for the induction of subsets of high osmolarity glycerol pathway-dependent genes. *Journal of Biological Chemistry*, **275:** 8290–8300.

REP, M., DUYVESTEIJN, R. G. E., GALE, L., USGAARD, T., CORNELISSEN, B. J. C., MA, L. J. & WARD, T. J. (2006) The presence of GC-AG introns in *Neurospora crassa* and other euascomycetes determined from analyses of complete genomes: Implications for automated gene prediction. *Genomics*, **87:** 338–347.

RICHARDS, T. A., LEONARD, G., SOANES, D. M. & TALBOT, N. J. (2011) Gene transfer in fungi. *Fungal Biology Reviews*, **25:** 98–110.

ROBBERTSE, B., REEVES, J. B., SCHOCH, C. L. & SPATAFORA, J. W. (2006) A phylogenetic analysis of the Ascomycota. *Fungal Genetics & Biology*, **43:** 715–725.

RODRIGUEZ-NAVARRO, A., BLATT, M. R. & SLAYMAN, C. L. (1986) A potassium proton-symport in *Neurospora crassa*. *The Journal of General Physiology*, **87:** 649–674.

ROKAS, A., PAYNE, G., FEDOROVA, N. D., BAKER, S. E., MACHIDA, M., YU, J., GEORGIANNA, D. R., DEAN, R. A., BHATNAGAR, D., CLEVELAND, T. E., WORTMAN, J. R., MAITI, R., JOARDAR, V., AMEDEO, P., DENNING, D. W. & NIERMAN, W. C. (2007) What can comparative genomics tell us about species concepts in the genus *Aspergillus*? *Studies in Mycology*, **59:** 11–17.

ROY, S. W. & GILBERT, W. (2006) The evolution of spliceosomal introns: patterns, puzzles and progress. *Nature Reviews Genetics*, **7:** 211–221.

ROY, S. W. & IRIMIA, M. (2009) Mystery of intron again: new data and new models. *Trends in Genetics*, **25:** 67–73.

RUIJTER, G. & VISSER, J. (1997) Carbon repression in *Aspergilli*. *FEMS Microbiology Letters*, **151:** 103–114.

RUIJTER, G., BAX, M., PATEL, H., FLITTER, S. J., VAN DE VONDERVOORT, P. J. I., DE VRIES, R. P., VANKUYK, P. A. & VISSER, J. (2003) Mannitol is required for stress tolerance in *Aspergillus niger* conidiospores. *Eukaryotic Cell*, **2:** 690–698.

RUIZ-HERRERA, J. (1994) Polyamines, DNA methylation, and fungal differentiation. *Critical Reviews in Microbiology*, **20:** 143–150.

SAITO, H. & TATEBAYASHI, K. (2004) Regulation of the osmoregulatory HOG MAPK cascade in yeast. *Journal of Biochemistry*, **136:** 267–272.

SALAZAR, M., VONGSANGNAK, W., PANAGIOTOU, G., ANDERSEN, M. R. & NIELSEN, J. (2009) Uncovering transcriptional regulation of glycerol metabolism in *Aspergilli* through genome-wide gene expression data analysis. *Molecular Genetics & Genomics*, **282:** 571–586.

SALOHEIMO, A., ARO, N., ILMÉN, M. & PENTTILÄ, M. (2000) Isolation of the *ace1* gene encoding a Cys2-His2 transcription factor involved in regulation of activity of the cellulase promoter *cbh1* of *Trichoderma reesei*. *Journal of Biological Chemistry*, **275:** 5817–5825.

SAMBROOK, J. & RUSSELL, D. W. (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.

SCHELLENBERG, M. J., RITCHIE, D. B. & MACMILLAN, A. M. (2008) Pre-mRNA splicing: a complex picture in higher definition. *Trends in Biochemical Sciences*, **33:** 243–246.

SELDEEN, K. L., McDONALD, C. B., DEEGAN, B. J. & FAROOQ, A. (2009) Single nucleotide variants of the TGACTCA motif modulate energetics and orientation of binding of the Jun-Fos heterodimeric transcription factor. *Biochemistry*, **48:** 1975–1983.

SEMIGHINI, C. P. GOLDMAN, G. H. & HARRIS, S. D. (2005) The DNA damage response of filamentous fungi: Novel features associated with a multicellular lifestyle. *Applied Mycology and Biotechnology,* **5:** 117–139.

SHEN, J. K. (2010) Uncovering specific electrostatic interactions in the denatured states of proteins. *Biophysical Journal*, **99:** 924–932.

SHERTZER, H. G., LÅSTBOM, L., SAINSBURY, M. & MOLDÉUS, P. Menadione-mediated membrane fluidity alterations and oxidative damage in rat hepatocytes (1992) *Biochemical Pharmacology,* **43:** 2135–2141.

SHI, J., CHEN, W., LIU, Q., CHEN, S., HU, H., TURNER, G. & LU, L. (2008) Depletion of the MobB and CotA complex in *Aspergillus nidulans* causes defects in polarity maintenance that can be suppressed by the environment stress. *Fungal Genetics & Biology*, **45:** 1570–1581.

SHRIVASTAV, M., DE HARO, L. P. & NICKOLOFF, J. A. (2008) Regulation of DNA double-strand break repair pathway choice. *Cell Research*, **18:** 134–147.

SIMS, A. H., GENT, M. E., ROBSON, G. D., DUNN-COLEMAN, N. S. & OLIVER, S. G. (2004) Combining transcriptome data with genomic and cDNA sequence alignments to make confident functional assignments for *Aspergillus nidulans* genes. *Mycological Research*, **108**: 853–857.

SLEATOR, R. D. (2010) An overview of the current status of eukaryotic gene prediction strategies. *Gene*, **461:** 1–4.

SMALL, A. J., HYNES, M. J. & DAVIS, M. A. (1999) The TamA protein fused to a DNA-binding domain can recruit AreA, the major nitrogen regulatory protein, to activate gene expression in *Aspergillus nidulans*. *Genetics*, **153:** 95–105.

SOLOMON, P. S., LEE, R. C., GREER-WILSON, T. J. & OLIVER, R. P. (2004) Pathogenicity of *Stagonospora nodorum* requires malate synthase. *Molecular Microbiology*, **53:** 1065–1073.

SOPHIANOPOULOU, V. & SCAZZOCCHIO, C. (1989). The proline transport protein of *Aspergillus nidulans* is very similar to amino acid transporters of *Saccharomyces cerevisiae. Molecular Microbiology*, **3:** 703–714.

SPATHAS, D. H. (1978) A salt sensitive mutation on chromosome VI of *Aspergillus nidulans*. *Aspergillus Newsletter*, **14:** 28

SPIELVOGEL, A., FINDON, H., ARST, H. N., ARAÚJO-BAZÁN, L., HERNÁNDEZ-ORTÍZ, P., STAHL, U., MEYER, V. & ESPESO, E. A. (2008) Two zinc finger transcription factors, CrzA and SltA, are involved in cation homeostasis and detoxification in *Aspergillus nidulans*. *Biochemical Journal*, **414:** 419–429.

STACK, M. & RODRICKS, J. V. (1971) Method for chemical confirmation of sterigmatocystin. *Association of Official Analytical Chemists*, **54**: 86–90.

STAJICH, J. E. & DIETRICH, F. S. (2006) Evidence of mRNA-mediated intron loss in the human-pathogenic fungus *Cryptococcus neoformans*. *Eukaryotic Cell*, **5:** 789–793.

STEIDL, S., PAPAGIANNOPOULOS, P., LITZKA, O., ANDRIANOPOULOS, A., DAVIS, M. A., BRAKHAGE, A. A. & HYNES, M. J. (1999) AnCF, the CCAAT binding complex of *Aspergillus nidulans*, contains products of the *hapB*, *hapC* and *hapE* genes and is required for activation by the pathway-specific regulatory gene *amdR*. *Molecular & Cellular Biology*, **19:** 99–106.

STRAUSS, J., MURO-PASTOR, M. I. & SCAZZOCCHIO, C. (1998) The regulator of nitrate assimilation in ascomycetes is a dimer which binds a nonrepeated, asymmetrical sequence. *Molecular & Cellular Biology*, **18:** 1339–1348.

STRAUSS, J., HORVATH, H. K., ABDALLAH, B. M., KINDERMANN, J., MACH, R. L. & KUBICEK, C. P. (1999) The function of CreA, the carbon catabolite repressor of *Aspergillus nidulans*, is regulated at the transcriptional and post-transcriptional level. *Molecular Microbiology*, **32:** 169–178.

SUZUKI, M. (1989) SPXX, a frequent sequence motif in gene regulatory proteins. *Journal of Molecular Biology*, **207:** 61–84.

SUZUKI, A., KANAMARU, K., AZUMA, N., KATO, M. & KOBAYASHI, T. (2008) GFP-tagged expression analysis revealed that some histidine kinases of *Aspergillus nidulans* show temporally and spatially different expression during the life cycle. *Bioscience, Biotechnology & Biochemistry*, **72:** 428–434.

SUZUKI, Y., MURRAY, S. L., WONG, K. H., DAVIS, M. A. & HYNES, M. J. (2012) Reprogramming of carbon metabolism by the transcriptional activators AcuK and AcuM in *Aspergillus nidulans*. *Molecular Microbiology,* **84:** 942–964.

SZEWCZYK, E., NAYAK, T., OAKLEY, C. E., EDGERTON, H., XIONG, Y., TALESH-TAHERI, N., OSMANI, S. A. & OAKLEY, B. (2006) Fusion PCR and gene targeting in *Aspergillus nidulans*. *Nature Protocols*, **1:** 3111–3120.

TADEO, X., LÓPEZ-MÉNDEZ, B., TRIGUEROS, T., LAÍN, A., CASTAÑO, D. & MILLET, O. (2009) Structural basis for the amino acid composition of proteins from halophilic archea. *PLoS Biology*, **7:** e1000257**.**

TAKAHASHI, T., MASUDA, T. & KOYAMA, Y. (2006) Enhanced gene targeting frequency in *ku70* and *ku80* disruption mutants of *Aspergillus sojae* and *Aspergillus oryzae*. *Molecular Genetics and Genomics*, **275:** 460–470.

TAMÁS, M. J., KARLGREN, S., BILL, R. M., HEDFALK, K., ALLEGRI, L., FERREIRA, M., THEVELEIN, J. M., RYDSTRÖM, J., MULLINS, J. G. & HOHMANN, S. (2003) A short regulatory domain restricts glycerol transport through yeast Fps1p. *Journal of Biological Chemistry*, **278:** 6337–6345.

TAMAYO, E. N., VILLANUEVA, A., HASPER, A. A., DE GRAAFF, L. H., RAMÓN, D. & OREJAS, M. (2008) CreA mediates repression of the regulatory gene *xlnR* which controls the production of xylanolytic enzymes in *Aspergillus nidulans*. *Fungal Genetics & Biology*, **45:** 984–993.

TATEBAYASHI, K., TANAKA, K., YANG, H. Y., YAMAMOTO, K., MATSUSHITA, Y., TOMIDA, T., IMAI, M. & SAITO, H. (2007) Transmembrane mucins Hkr1 and Msb2 are putative osmosensors in the SHO1 branch of yeast HOG pathway. *EMBO Journal*, **26:** 3521–3533.

THÖN, M., AL ABDALLAH, Q., HORTSCHANSKY, P., SCHARF, D. H., EISENDLE, M., HAAS, H. & BRAKHAGE, A. A. (2010) The CCAAT-binding complex co-ordinates the oxidative stress response in eukaryotes. *Nucleic Acids Research*, **38:** 1098–1113.

TILBURN, J., SCAZZOCCHIO, C., TAYLOR, G. G., ZABICKY-ZISSMAN, J. H., LOCKINGTON, R. A. & DAVIES, R. W. (1983) Transformation by integration in *Aspergillus nidulans*. *Gene*, **26:** 205–221.

TILBURN, J., SARKAR, S., WIDDICK, D. A., ESPESO, E. A., OREJAS, M., MUNGROO, J., PEÑALVA, M. A. & ARST, H. N. (1995) The *Aspergillus* PacC zinc finger transcription factor mediates regulation of both acid- and alkaline-expressed genes by ambient pH. The *EMBO Journal*, **14:** 779–790.

TIROSH, I., WEINBERGER, A., BEZALEL, D., KAGANOVICH, M. & BARKAI, N. (2008) On the relation between promoter divergence and gene expression evolution. *Molecular Systems Biology*, **4:** 159.

TODD, R. B., ANDRIANOPOULOS, A., DAVIS, M. A. & HYNES, M. J. (1998) FacB, the *Aspergillus nidulans* activator of acetate utilization genes, binds dissimilar DNA sequences. *EMBO Journal*, **17:** 2042–2054.

Tree of Life Web Project (2006) Pezizomycotina (http://tolweb.org/Pezizomycotina/29296). Version 9, October 2006.

TSAI, H. K., HUANG, G. T. W., CHOU, M. Y., LU, H. H. S. & LI, W. H. (2006) Method for identifying transcription factor binding sites in yeast. *Bioinformatics*, **22:** 1675–1681.

VAN HEEMST, D., SWART, K., HOLUB, E. F., VAN DIJK, R., OFFENBERG, H. H., GOOSEN, T., VAN DEN BROEK, H. W. & HEYTING, C. (1997) Cloning, sequencing, disruption and phenotypic analysis of *uvsC*, an *Aspergillus nidulans* homologue of yeast *RAD51*. *Molecular & General Genetics*, **254:** 654–664.

VIRGINIA, M., APPLEYARD, C. L., McPHEAT, W. L. & STARK, M. J. R. (2000) A novel 'two-component' protein containing histidine kinase and response regulator domains required for sporulation in *Aspergillus nidulans*. *Current Genetics*, **37:** 364–372.

WAGNER, G. P. & LYNCH, V. J. (2008) The gene regulatory logic of transcription factor evolution. *Trends in Ecology & Evolution*, **23:** 377–385.

WAKABAYASHI, M., ISHII, C., HATAKEYAMA, S., INOUE, H. & TANAKA, S. (2010) ATM and ATR homologues of *Neurospora crassa* are essential for normal cell growth and maintenance of chromosome integrity. *Fungal Genetics & Biology*, **47:** 809–817.

WALTERS, D. R., COWLEY, T. & McPHERSON, A. (1997) Polyamine metabolism in the thermolerant mesophilic fungus *Aspergillus fumigatus*. *FEMS Microbiology Letters*, **153:** 433–437.

WANG, H., XU, Z., GAO, L. & HAO, B. (2009) A fungal phylogeny based on 82 complete genomes using the composition vector method. *BMC Evolutionary Biol*ogy, **9:** 195.

WARRIS, A. & VERWEIJ, P. E. (2005) Clinical implications of environmental sources for *Aspergillus*. *Medical Mycology*, **43:** S59–S65.

WASLYNKA, J. A. & MOORE, M. M. (2003) *Aspergillus fumigatus* conidia survive and germinate in acidic organelles of A549 epithelial cells. *Journal of Cell Science*, **116:** 1579–1587.

WHITEHEAD, M. P. (1990) *Development of homologous transformation systems for the filamentous fungi Cephalosporium acremonium and Penicillium chrysogenum*. PhD thesis, University of St. Andrews.

WILSON, R. A. & ARST, H. N. (1998) Mutational analysis of AreA, a transcriptional activator mediating nitrogen metabolite repression in *Aspergillus nidulans* and a member of the 'streetwise' GATA family of transcription factors. *Microbiology & Molecular Biology Reviews*, **62:** 586–596.

WITTCOPP, P. J. (2010) Variable transcription factor binding: a mechanism of evolutionary change. *PLoS Biology*, **8:** e1000342.

WITTEVEEN, C. F. B. & VISSER, J. (1995) Polyol pools in *Aspergillus niger*. *FEMS Microbiology Letters*, **134:** 57–62.

WONG, K. H., HYNES, M. J., TODD, R. B. & DAVIS, M. A. (2007) Transcriptional control of *nmrA* by the bZIP transcription factor MeaB reveals a new level of nitrogen regulation in *Aspergillus nidulans*. *Molecular Microbiology*, **66:** 534–551.

WONG, K. H., HYNES, M. J., TODD, R. B. & DAVIS, M. A. (2009) Deletion and overexpression of the *Aspergillus nidulans* GATA factor AreB reveals unexpected pleiotropy. *Microbiology*, **155:** 3868–3880.

WOODCOCK, N. A. (1997) *Biochemical and physiological responses of Aspergillus nidulans to osmotic stress*. Ph.D. Thesis, University of Wolverhampton.

WORTMAN, J. R., FEDOROVA, N., CRABTREE, J., JOARDER, V., MAITI, R., HAAS, B. J., AMEDEO, P., LEE, E., ANGIUOLI, S. V., JIANG, B., ANDERSEN, M. J., DENNING, D. W., WHITE, O. R. & NIERMAN, W. C. (2006) Whole genome comparison of the *A. fumigatus* family. *Medical Mycology*, **44:** S3–S7.

WORTMAN, J. R., GILSENAN, J. M., JOARDAR, V., DEEGAN, J., CLUTTERBUCK, A. J., ANDERSEN, M. R., ARCHER, D., BENCINA, M., BRAUS, G., COUTINHO, P., VON DÖHREN, H., DOONAN, J., DRIESSEN, A. J. M., DUREK, P., ESPESO, E., FEKETE, E., FLIPPHI, M., ESTRADA, C. G., GEYSENS, S., GOLDMAN, G., DE GROOT, P. W. J., HANSEN, K., HARRIS, S. D., HEINEKAMP, T., HELMSTAEDT, K., HENRISSAT, B., HOFMANN, G., HOMAN, T., HORIO, T., HORIUCHI, H., JAMES, S., JONES, M., KARAFFA, L., KARÁNYI, Z., KATO, M., KELLER, N., KELLY, D. E., KIELY, J. A. K. W., KIM, J. M., VAN DER KLEI, I. J., KLIS, F. M., KOVALCHUK, A., KRASEVEC, N., KUBICEK, C. P., LIU, B., MACCABE, A., MEYER, V., MIRABITO, P., MISKEI, M., MOS, M., MULLINS, J., NELSON, D. R., NIELSEN, J., OAKLEY, B. R., OSMANI, S. A., PAKULA, T., PASZEWSKI, A., PAULSEN, I., PILSYK, S., PÓCSI, I., PUNT, P. J., RAM, A. F. J., REN, Q., ROBELLET, X., ROBSON, G., SEIBOTH, B., VAN SOLINGEN, P., SPECHT, T., SUN, J., TAHERI-TALESH, N., TAKESHITA, N., USSERY, D., VANKUYK, P. A., VISSER, H., VAN DE VONDERVOORT, P. J. I., DE VRIES, R. P., WALTON, J., XIANG, X., XIONG, Y., ZENG, A. P., BRANDT, B. W., CORNELL, M. J., VAN DEN HONDEL, C. A. M. J., VISSER, J., OLIVER, S. G. & TURNER, G. (2009) The 2008 update of the *Aspergillus nidulans* genome annotation: A community effort. *Fungal Genetics & Biology*, **46:** S2–S13.

WYATT, M. D. & PITTMANN, D. L. (2006) Methylating agents and DNA repair responses: Methylated bases and sources of strand breaks. *Chemical Research in Toxicology*, **19:** 1580–1594.

YANG, L., UKIL, L., OSMANI, A., NAHM, F., DAVIES, J. DE SOUZA, C. P. C., DOU, X., PEREZ-BALAGUER, A. & OSMANI, S. A. (2004) Rapid production of gene replacement constructs and generation of

a green fluorescent protein-tagged centromeric marker in *Aspergillus nidulans*. *Eukaryotic Cell*, **3:** 1359–1362.

YU, J., CHANG, P. K., EHRLICH, K. C., CARY, J. W., BHATNAGAR, D., CLEVELAND, T. E., PAYNE, G. A., LINZ, J. E., WOLOSHUK, C. P. & BENNETT, J. W. (2004a) Clustered pathway genes in aflatoxin biosynthesis. *Applied & Environmental Microbiology*, **70:** 1253–1262.

YU, J. H., HAMARI, Z., HAN, K. H., SEO, J. A., REYES-DOMINGUEZ, Y. & SCAZZOCCHIO, C. (2004b) Double-joint PCR: a PCR-based molecular tool for gene manipulations in filamentous fungi. *Fungal Genetics & Biology*, **41:** 973–981.

ZHANG, Z. & TOWNSEND, J. P. (2010) The filamentous fungal gene expression database (FFGED). *Fungal Genetics & Biology*, **47:** 199–204.

ZHAO, Q., YUAN, S., ZHANG, Y., ZHU, H., DAI, C., YANG, F. & HAN, F. (2007) Expression, purification and characterization of pectate lyase A from *Aspergillus nidulans* in *Escherichia coli*. *World Journal of Microbiology and Biotechnology*, **23:** 1057–1064.

ZHENG, W., ZHAO, H., MANCERA, E., STEINMETZ, L. M. & SNYDER, M. (2010) Genetic analysis of variation in transcription factor binding in yeast. *Nature*, **464:** 1187–1191.